



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 163760

TO: Patricia Duffy
Location: 3b05 / 3c18
Sunday, August 28, 2005
Art Unit: 1645
Phone: 571-272-0855
Serial Number: 10 / 078531

From: Jan Delaval
Location: Biotech-Chem Library
Remsen 1a51
Phone: 571-272-2504
jan.delaval@uspto.gov

Search Notes

THIS PAGE BLANK (USPTO)

From: Duffy, Patricia
Sent: Thursday, August 25, 2005 8:37 AM
To: STIC-Biotech/ChemLib
Subject: sequence search

In re: 10/078,531

Please search SEQ ID NO:2 and oligomers thereof.
Please include both a sequence and interference search.
Please print out top 100 hits in each category.

Thank you.

Patricia A. Duffy, Ph.D.
Art Unit 1645
Remsen 3B05; Mailbox 3C18
571-272-0855

STAFF USE ONLY

Searcher: [Signature]
Searcher Phone: 2- 2504
Date Searcher Picked up: 8/28/05
Date Completed: 8/28/05
Searcher Prep/Rev. Time: 20
Online Time: 15

Type of Search

NA#: _____ AA#: 2
Interference: ✓ SPDI: _____
S/L: _____ Oligomer: ✓
Encode/Transl: _____
Structure#: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: ✓
WWW/Internet: _____
Other(Specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:28:40 ; Search time 182 Seconds
(without alignments)
2836.131 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 5080

Sequence: 1 MKKHUKTVALTLTVSVVTH.....LGYTSVALSLTAIKKKY 1008

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 110 summaries

Database : UniProt_03.*

1: uniprot_spot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5080	100.0	1008	2	Q9A0C0
2	5058	99.6	1008	2	Q8P1F7
3	5040	99.2	1008	2	Q878I7
4	4993	98.3	999	2	Q87XJ8
5	3776.5	74.3	1055	2	Q8E1D6
6	3771.5	74.2	1055	2	Q8B6V3
7	331.5	6.5	1081	2	Q68831
8	317	6.2	1395	2	Q8T164
9	294.5	5.8	1518	2	Q73M50
10	263	5.2	1734	2	Q8T167
11	247	4.9	1066	2	Q7PG02
12	247	4.9	1743	2	Q8T171
13	247	4.9	4734	2	Q74JH9
14	230.5	4.5	1139	2	Q81JS7
15	229	4.5	3317	2	Q8SWP8
16	227	4.5	1523	2	Q7RS28
17	225	4.4	178	2	Q8GLH0
18	224.5	4.4	2719	2	Q7PG02
19	223	4.4	2664	2	Q7REU0
20	221	4.4	690	2	Q8DN38
21	220.5	4.3	3257	2	Q9V736
22	220	4.3	1443	2	Q9VJQ0
23	220	4.3	1471	2	Q7KW92
24	219.5	4.3	1828	2	Q700S8
25	219	4.3	1208	1	PCP1_SCHPO
26	218.5	4.3	1675	2	Q7RHM8
27	218	4.3	662	2	Q6K1F4
28	218	4.3	2341	2	Q7RR45
29	217.5	4.3	1072	2	Q9CF64
30	217.5	4.3	1372	2	Q66275
31	217.5	4.3	1372	2	Q66283

005166 streptococc
Q7RMS4 plasmodium
Q14789 homo sapien
Q97NB5 streptococc
Q71WB3 listeria mo
Q7XN10 oryza sativ
Q7RGP2 plasmodium
Q833V2 enterococcu
Q9AOK5 streptococc
Q7RFS2 plasmodium
Q7PRD4 anopheles g
Q9GZ9 mycoplasma
Q7REY3 plasmodium
Q6HBX5 bacillus th
Q6BY65 debaryomyce
Q07910 helicobacte
Q9AHK6 borrelia bu
Q9BJY0 plasmodium
Q97K41 clostridium
Q00799 plasmodium
Q6CPF6 kluyveromyc
Q8S9G0 clostridium
Q00798 plasmodium
Q8RGS5 fusobacteri
Q64VN1 bacteroides
Q839L8 enterococcu
Q845L8 mycoplasma
Q26223 plasmodium
Q750B4 ashbya goss
Q8CP76 staphylococ
Q73CV3 bacillus ce
Q99R6 staphylococ
Q74B1 staphylococ
Q92H10 haemophilus
Q7RTL4 drosophila
Q01385 drosophila
Q8X6A0 fusobacteri
Q72X91 bacillus ce
Q931R6 staphylococ
Q99U54 staphylococ
Q723X5 listeria mo
Q7Q168 anopheles g
Q869B8 dictyosteli
Q9AV25 oryza sativ
Q7WRQ2 staphylococ
Q7RNW8 plasmodium
Q7YWF0 plasmodium
Q6G9X3 staphylococ
Q6YA77 plasmodium
Q6BUQ9 debaryomyce
Q81K49 plasmodium
Q64046 bacterioph
Q31976 bacillus su
Q7YWE8 plasmodium
Q99YA0 streptococc
Q7Q8A9 anopheles g
Q7WE91 staphylococ
Q99TB0 staphylococ
Q6GFU5 staphylococ
Q8IEB6 plasmodium
Q8NWQ6 staphylococ
P25386 saccharomyc
Q9V624 drosophila
Q9EXH7 listeria iv
Q86ZA8 cochllobolu
Q07380 saccharomyc
Q9U4X0 plasmodium
Q7P2P7 fusobacteri
Q8GC27 listeria iv
Q7B120 borrelia bu
Q9YV76 melanioplus
Q8NZK6 streptococc
Q814R4 bacillus ce

105 197 3.9 2441 2 096124 O96124 plasmodium
106 197 3.9 4007 2 075H24 Q75H24 neurospora
107 196.5 3.9 1680 2 09U8G1 Q9U8G1 plasmodium
108 196.5 3.9 1875 1 MLF1_YEAST Q04551 saccharomyc
109 196.5 3.9 2139 2 Q07569 Q07569 entamoeba h
110 196.5 3.9 2479 2 Q66GS8 Q66GS8 homo sapien

ALIGNMENTS

RESULT 1
Q9A0C0
ID Q9A0C0 PRELIMINARY; PRT; 1008 AA.
AC Q9A0C0
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Hypothetical protein SPY0843.
GN OrderedLocNames=SPY0843;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Lyon K.,
RA Primeaux C., Sezzate S., Suvoarov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
DR EMBL; AB006534; AAK33772.1; -
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR007093; LRR_Tp.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1008 AA; 111503 MW; 045BF8CE931AF0CF CRC64;

Query Match 100.0%; Score 5080; DB 2; Length 1008;
Best Local Similarity 100.0%; Pred. No. 2.4e-215; Indels 0; Gaps 0;
Matches 1008; Conservative 0; Mismatches 0;

QY 1 MKKHLKTLVALTLTVSVTHNQEVFSLVKEPIKQTSASSISGADYAESGSKLKINE 60
DB 1 MKKHLKTLVALTLTVSVTHNQEVFSLVKEPIKQTSASSISGADYAESGSKLKINE 60

QY 61 TSGPVDVTVDLPSDKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120
DB 61 TSGPVDVTVDLPSDKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120

QY 121 SLSLNTKVPSTGNWEICDITKGNLTGLVSKSGVKLSOTDHLVLPQAADGTQLIQVAS 180
DB 121 SLSLNTKVPSTGNWEICDITKGNLTGLVSKSGVKLSOTDHLVLPQAADGTQLIQVAS 180

QY 181 FAFTPDKKTAAIAYTSRAGENGESQLDVGKEIINEGEVFNLSYLKKVTTPTGYKHIGQ 240
DB 181 FAFTPDKKTAAIAYTSRAGENGESQLDVGKEIINEGEVFNLSYLKKVTTPTGYKHIGQ 240

QY 241 DAFVFNKNTAFVNLPSLETISDYAFALHAKQIDLPNLKAIGLAFPDNQITKLSLP 300
DB 241 DAFVFNKNTAFVNLPSLETISDYAFALHAKQIDLPNLKAIGLAFPDNQITKLSLP 300

QY 301 ROLMLAERAFKSNHKTTEFGNSLKVIGEASFQNDLSQLMLPDGLEKIESAFTGNP 360
DB 301 ROLMLAERAFKSNHKTTEFGNSLKVIGEASFQNDLSQLMLPDGLEKIESAFTGNP 360

QY 361 GDDHYNNRVVLTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB 361 GDDHYNNRVVLTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

QY 421 GFSNKGLOKVRKNKLEIPKOHNGVTITTEIGDNAFRNVDFONKTLRKVDLEEVKLPSTIR 480
DB 421 GFSNKGLOKVRKNKLEIPKOHNGVTITTEIGDNAFRNVDFONKTLRKVDLEEVKLPSTIR 480

QY 481 KIGAFAFQSNLKSFEASDDLEETKEGAFMNNRITETLELKDVLVTIGDAAPHNIIYAIV 540
DB 481 KIGAFAFQSNLKSFEASDDLEETKEGAFMNNRITETLELKDVLVTIGDAAPHNIIYAIV 540

QY 541 LPESVQIEGRSAFRONGANNLI FMGSKVKTIGEMAFISNRLEHLDLSEKQLTPIVQAF 600
DB 541 LPESVQIEGRSAFRONGANNLI FMGSKVKTIGEMAFISNRLEHLDLSEKQLTPIVQAF 600

QY 601 SDNALKEVLLPASLKTITREAFKKNHLKOLEVASALSASHIENALDDNDGDGDFONKVVVK 660
DB 601 SDNALKEVLLPASLKTITREAFKKNHLKOLEVASALSASHIENALDDNDGDGDFONKVVVK 660

QY 661 THNSYALADGEHFIIVDPDKLSSTIVLEKILKIEGLDYSTLRQTQTQTFQFRDMTTAGKA 720
DB 661 THNSYALADGEHFIIVDPDKLSSTIVLEKILKIEGLDYSTLRQTQTQTFQFRDMTTAGKA 720

QY 721 LLSKSNLRQGEKQFLQEAQFPLGRVLDLKAIAKAERKALVTKKATQNGQLLERSINKAVL 780
DB 721 LLSKSNLRQGEKQFLQEAQFPLGRVLDLKAIAKAERKALVTKKATQNGQLLERSINKAVL 780

QY 781 AYNSAISKANVKELEKELDLITGLVEGKGLAQATWVGYYLTKTLPPLPYVYIGLVNY 840
DB 781 AYNSAISKANVKELEKELDLITGLVEGKGLAQATWVGYYLTKTLPPLPYVYIGLVNY 840

QY 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDDNEGYSALAVATLADYEGLDIKITILN 900
DB 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDDNEGYSALAVATLADYEGLDIKITILN 900

QY 901 SKLSQLTISROVPTAAVHRAGIFQAIQNAABAEQQLPKPOTHSEKSSSSSSANSKORGL 960
DB 901 SKLSQLTISROVPTAAVHRAGIFQAIQNAABAEQQLPKPOTHSEKSSSSSSANSKORGL 960

QY 961 QSNPKTNRGRHSAILPRTGSKGSPVYGLVTSVALLSLITAIKKKY 1008
DB 961 QSNPKTNRGRHSAILPRTGSKGSPVYGLVTSVALLSLITAIKKKY 1008

RESULT 2
QSP1F7
ID Q8P1F7 PRELIMINARY; PRT; 1008 AA.
AC Q8P1F7;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Hypothetical protein spyM18_0903.
GN OrderedLocNames=spyM18_0903;
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232;
RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.0625260099;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.B., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Mueser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AB010019; AAL97553.1; -
DR GO; GO:0009986; C:cell surface; IEA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR007093; LRR_Tp.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Complete proteome; Hypothetical protein.

SQ	SEQUENCE	1008 AA; 111620 MW; 89E979683FC2FF03B CRC64;
Query Match		
Best Local Similarity 99.6%; Score 5058; DB 2; Length 1008;		
Matches 1004; Conservative 1; Mismatches 3; Indels 0; Gaps 0;		
QY	1	MKKHLKTVLTLTTVVVTHNQEVSLVKEPILKQTOASSISGADYAESGSKLKINE 60
DB	1	MKKHLKTVLTLTTVVVTHNQEVSLVKEPILKQTOASSISGADYAESGSKLKINE 60
QY	61	TSGPVDVTVDLFSKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120
DB	61	TSGPVDVTVDLFSKRTTPEKIKONLAKPREQELKAVTENSEKQINSGSQLEQSK 120
QY	121	SLSLNTKVPSTSNWEICDFITKGNLTGLVLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
DB	121	SLSLNTKVPSTSNWEICDFITKGNLTGLVLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
QY	181	FAFTPDKKTALAEYTSRAGENGESIQDLDVKGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
DB	181	FAFTPDKKTALAEYTSRAGENGESIQDLDVKGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
QY	241	DAFVNDKNIAEVLNPESLETISDYAFALHAKQIDLPNLKKAIGELAFDNOITGKLSLP 300
DB	241	DAFVNDKNIAEVLNPESLETISDYAFALHAKQIDLPNLKKAIGELAFDNOITGKLSLP 300
QY	301	QRLMLAEARAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360
DB	301	QRLMLAEARAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360
QY	361	GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB	361	GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
QY	421	PFAM; PF00746; Gram_pos_anchor; 1.
DB	421	PFAM; PF00746; Gram_pos_anchor; 1.
QY	481	KIDAFQSNLKSFEASDDLEIEKEGAFMNNRIETLEKDKLVITIGDAAFHINHIYAIV 540
DB	481	KIDAFQSNLKSFEASDDLEIEKEGAFMNNRIETLEKDKLVITIGDAAFHINHIYAIV 540
QY	541	LPBSVOEIGRSFRQNGANNLIFMGSKVKTLGEMAFNLNRLHDLSEKQQLTEIPVQAF 600
DB	541	LPBSVOEIGRSFRQNGANNLIFMGSKVKTLGEMAFNLNRLHDLSEKQQLTEIPVQAF 600
QY	601	SDNALKEVLLPASLTIRIEEAFKQHLKQLEVASALSHIAFNALDNDGDGEQFNKVVVK 660
DB	601	SDNALKEVLLPASLTIRIEEAFKQHLKQLEVASALSHIAFNALDNDGDGEQFNKVVVK 660
QY	661	THNSYALADGSHFIVDPDKLSSTIVDLKILKEGLDYSTRLOTOTOPRDMTTAGKA 720
DB	661	THNSYALADGSHFIVDPDKLSSTIVDLKILKEGLDYSTRLOTOTOPRDMTTAGKA 720
QY	721	LLSKNLRQGEKQKLOEAPFLGRVDLDKATAEAKALVTKATKNGQLLERSINKAVL 780
DB	721	LLSKNLRQGEKQKLOEAPFLGRVDLDKATAEAKALVTKATKNGQLLERSINKAVL 780
QY	781	AYNNSAIKANKVRLEKELDLTLGLVEGKPLAQATWQGVYLLKTPPLPBYIYGLNVY 840
DB	781	AYNNSAIKANKVRLEKELDLTLGLVEGKPLAQATWQGVYLLKTPPLPBYIYGLNVY 840
QY	841	PDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGVHALAVATLADYEGLDIKTILN 900
DB	841	PDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGVHALAVATLADYEGLDIKTILN 900
QY	901	SKLSQTSIRQVPTAAYTHRAGIFQAIQNAAAEAEQLPKPGTHSEKSSSSSESANSKORGL 960
DB	901	SKLSQTSIRQVPTAAYTHRAGIFQAIQNAAAEAEQLPKPGTHSEKSSSSSESANSKORGL 960
QY	961	QSNPKTNRGRSAILPRTGSGSGSFYVIGLYTSVALLSLIITAIIKKKY 1008
DB	961	QSNPKTNRGRSAILPRTGSGSGSFYVIGLYTSVALLSLIITAIIKKKY 1008

RESULT 3

QY	Q87817	PRELIMINARY; PRT; 1008 AA.
AC	Q87817	
DT	01-JUN-2003 (Tremblrel. 24, Created)	
DT	01-JUN-2003 (Tremblrel. 24, Last sequence update)	
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)	
DE	Hypothetical protein SPa1285.	
GN	OrderedLocusNames=SPa1285;	
OS	Streptococcus pyogenes (serotype M3).	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OC	Streptococcus.	
OX	NCBI TaxID=198466;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=SSI-1;	
RX	MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;	
RA	Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,	
RA	Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,	
RA	Hayaishi H., Hattori M., Hamada S.;	
RT	"Genome sequence of an M3 strain of Streptococcus pyogenes reveals a	
RT	large-scale genomic rearrangement in invasive strains and new insights	
RT	into phage evolution."	
RL	Genome Res. 13:1042-1055 (2003).	
CC	-!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by	
CC	an amide bond (By similarity).	
CC	EMBL; AP005145; BAC64380.1; -.	
DR	GO; GO:0009986; C:cell surface; IEA.	
DR	GO; GO:0005618; C:cell wall; IEA.	
DR	InterPro; IPR001899; Gram_pos_anchor.	
DR	InterPro; IPR007093; LRR_Tp.	
DR	Pfam; PF00746; Gram_pos_anchor; 1.	
DR	TIGRPFAM; TIGR01167; LPATG_anchor; 1.	
DR	PROSITE; PS00847; GRAM_POS_ANCHORING; 1.	
KW	Cell wall; Hypothetical protein; Peptidoglycan-anchor.	
SQ	SEQUENCE 1008 AA; 111537 MW; 793C598CE094EE92 CRC64;	
Query Match 99.2%; Score 5040; DB 2; Length 1008;		
Best Local Similarity 99.2%; Pred. No. 1.4e-21;		
Matches 1000; Conservative 3; Mismatches 5; Indels 0; Gaps 0;		
QY	1	MKKHLKTVLTLTTVVVTHNQEVSLVKEPILKQTOASSISGADYAESGSKLKINE 60
DB	1	MKKHLKTVLTLTTVVVTHNQEVSLVKEPILKQTOASSISGADYAESGSKLKINE 60
QY	61	TSGPVDVTVDLFSKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120
DB	61	TSGPVDVTVDLFSKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120
QY	121	SLSLNTKVPSTSNWEICDFITKGNLTGLVLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
DB	121	SLSLNTKVPSTSNWEICDFITKGNLTGLVLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
QY	181	FAFTPDKKTALAEYTSRAGENGESIQDLDVKGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
DB	181	FAFTPDKKTALAEYTSRAGENGESIQDLDVKGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
QY	241	DAFVNDKNIAEVLNPESLETISDYAFALHAKQIDLPNLKKAIGELAFDNOITGKLSLP 300
DB	241	DAFVNDKNIAEVLNPESLETISDYAFALHAKQIDLPNLKKAIGELAFDNOITGKLSLP 300
QY	301	QRLMLAEARAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360
DB	301	QRLMLAEARAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360
QY	361	GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB	361	GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
QY	421	GFSNKGLOKVRKNLEIPKQNGVTIIEIGNAFRNVDFQNTLRKYDLEEVKLPSTIR 480
DB	421	GFSNKGLOKVRKNLEIPKQNGVTIIEIGNAFRNVDFQNTLRKYDLEEVKLPSTIR 480
QY	481	KIDAFQSNLKSFEASDDLEIEKEGAFMNNRIETLEKDKLVITIGDAAFHINHIYAIV 540
DB	481	KIDAFQSNLKSFEASDDLEIEKEGAFMNNRIETLEKDKLVITIGDAAFHINHIYAIV 540
QY	541	LPBSVOEIGRSFRQNGANNLIFMGSKVKTLGEMAFNLNRLHDLSEKQQLTEIPVQAF 600
DB	541	LPBSVOEIGRSFRQNGANNLIFMGSKVKTLGEMAFNLNRLHDLSEKQQLTEIPVQAF 600
QY	601	SDNALKEVLLPASLTIRIEEAFKQHLKQLEVASALSHIAFNALDNDGDGEQFNKVVVK 660
DB	601	SDNALKEVLLPASLTIRIEEAFKQHLKQLEVASALSHIAFNALDNDGDGEQFNKVVVK 660
QY	661	THNSYALADGSHFIVDPDKLSSTIVDLKILKEGLDYSTRLOTOTOPRDMTTAGKA 720
DB	661	THNSYALADGSHFIVDPDKLSSTIVDLKILKEGLDYSTRLOTOTOPRDMTTAGKA 720
QY	721	LLSKNLRQGEKQKLOEAPFLGRVDLDKATAEAKALVTKATKNGQLLERSINKAVL 780
DB	721	LLSKNLRQGEKQKLOEAPFLGRVDLDKATAEAKALVTKATKNGQLLERSINKAVL 780
QY	781	AYNNSAIKANKVRLEKELDLTLGLVEGKPLAQATWQGVYLLKTPPLPBYIYGLNVY 840
DB	781	AYNNSAIKANKVRLEKELDLTLGLVEGKPLAQATWQGVYLLKTPPLPBYIYGLNVY 840
QY	841	PDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGVHALAVATLADYEGLDIKTILN 900
DB	841	PDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGVHALAVATLADYEGLDIKTILN 900
QY	901	SKLSQTSIRQVPTAAYTHRAGIFQAIQNAAAEAEQLPKPGTHSEKSSSSSESANSKORGL 960
DB	901	SKLSQTSIRQVPTAAYTHRAGIFQAIQNAAAEAEQLPKPGTHSEKSSSSSESANSKORGL 960
QY	961	QSNPKTNRGRSAILPRTGSGSGSFYVIGLYTSVALLSLIITAIIKKKY 1008
DB	961	QSNPKTNRGRSAILPRTGSGSGSFYVIGLYTSVALLSLIITAIIKKKY 1008

Db 421 GPSSKGLQKVKENKNLEIPKQHNGVTIITEIGDNAFRNVDFQNKTLRKLYDLEVKLPSTIR 480
Qy 481 KIGAFAPQSNLNKSPASDDLEIEIEKEGAFPMNRIETLEBKOKLVITIGDAAFHINHIYAIV 540
Db 481 KIGAFAPQSNLNKSPASDDLEIEIEKEGAFPMNRIETLEBKOKLVITIGDAAFHINHIYAIV 540
Qy 541 LPESVOETGRSAFRONGANNLIFMGSKVKTIGEMAFLSNRLEHLDLSEKQLTEIPVOAF 600
Db 541 LPESVOETGRSAFRONGANNLIFMGSKVKTIGEMAFLSNRLEHLDLSEKQLTEIPVOAF 600
Qy 601 SDNALKEVLLPASLKTIREEAFKQNLKQLEVASALSHTAFNALDDNDGDEOFNKVVVK 660
Db 601 SDNALKEVLLPASLKTIREEAFKQNLKQLEVASALSHTAFNALDDNDGDEOFNKVVVK 660
Qy 661 THNSYALADGHEFTVDPDKLSTSTVLDLEKILKEGLDYSTLRQTQTQFRDMTTAGKA 720
Db 661 THNSYALADGHEFTVDPDKLSTSTVLDLEKILKEGLDYSTLRQTQTQFRDMTTAGKA 720
Qy 721 LLSKSNLRQGEKQKQLEQAOFLGRVLDLKAIAKAERKALVTKATNGQLLERSINKAVL 780
Db 721 LLSKSNLRQGEKQKQLEQAOFLGRVLDLKAIAKAERKALVTKATNGQLLERSINKAVL 780
Qy 781 AYNSAIKANKVRLEKELDLITGLVEGKGPLAQATMWQVYLLKTPPLPEYIYGLNVY 840
Db 781 AYNSAIKANKVRLEKELDLITGLVEGKGPLAQATMWQVYLLKTPPLPEYIYGLNVY 840
Qy 841 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLDIKTILN 900
Db 841 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLDIKTILN 900
Qy 901 SKLSQLTSTIRQVPTAAYHRAGIFQAIQNAABAEQLLPKPGTHSEKSSSESANSKORGL 960
Db 901 SKLSQLTSTIRQVPTAAYHRAGIFQAIQNAABAEQLLPKPGTHSEKSSSESANSKORGL 960
Qy 961 QSNPNTNGRHSAILPRTGSGKSFVYGILGYTSVALLSLITAIKKKY 1008
Db 961 QSNPNTNGRHSAILPRTGSGKSFVYGILGYTSVALLSLITAIKKKY 1008

RESULT 4
Q8K7X8
ID Q8K7X8 PRELIMINARY; PRT; 999 AA.
AC Q8K7X8;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Putative surface antigen.
GN OrderedLocusNames=SpvM3_0569;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.,
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence."
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (By similarity).
DR EMBL; AE014148; AM9176.1; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR InterPro; IPR001899; C:cell wall; IEA.
DR InterPro; IPR007093; LRR Tp.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.

DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Complete proteome; Peptidoglycan-anchor.
SQ SEQUENCE 999 AA; 110518 MW; 39F9801882BAEF7C CRC64;
Query Match 98.3%; Score 4993; DB 2; Length 999;
Matches 990; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Qy 10 LTLTTSVVTHNQVFSVLKPEILKQTOQASSISGADYAESGSKLKNETSQVDDTV 69
Db 1 MTLTTSVVTHNQVFSVLKPEILKQTOQASSISGADYAESGSKLKNETSQVDDTV 60
Qy 70 TDLFSDKRTTPEKTKDNLAGPREQLKAVTENTSEKQITSGSQLESLSINKTVP 129
Db 61 TDLFSDKRTTPEKTKDNLAGPREQLKAVTENTSEKQITSGSQLESLSINKRVP 120
Qy 130 STSNWEICDFITTKGNTLVGLSKSGVEKLSQTDHLVPSQAADGTQLIQVASFATPDKKT 189
Db 121 STSNWEICDFITTKGNTLVGLSKSGVEKLSQTDHLVPSQAADGTQLIQVASFATPDKKT 180
Qy 190 AIAEYTSRAGENGHSIQDLDVKGKIIINEGEVFNYSLLKKVTIPTGYKHIGODAFVDNKVI 249
Db 181 AIAEYTSRAGENGHSIQDLDVKGKIIINEGEVFNYSLLKKVTIPTGYKHIGODAFVDNKVI 240
Qy 250 AEVNLPSLETSIDYAFALHALKQIDLPDNLKAIGELAFPDNOITGKLSLPRQLMLAER 309
Db 241 AEVNLPSLETSIDYAFALHALKQIDLPDNLKAIGELAFPDNOITGKLSLPRQLMLAER 300
Qy 310 AFKSNHKTTEFRGNLSKVIGEASFQNDLSQMLPDGLEKIESEAFTEGNGPDHNNRV 369
Db 301 AFKSNHKTTEFRGNLSKVIGEASFQNDLSQMLPDGLEKIESEAFTEGNGPDHNNRV 360
Qy 370 VLMTKSGKNPGLATENTYVNPDKSLMQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOK 429
Db 361 VLMTKSGKNPGLATENTYVNPDKSLMQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOK 420
Qy 430 VKRNKNIETPKQHNGVTITEIGDNAFRNVDFQNKTLRKLYDLEVKLPSTIRKIGAFAPQS 489
Db 421 VKRNKNIETPKQHNGVTITEIGDNAFRNVDFQNKTLRKLYDLEVKLPSTIRKIGAFAPQS 480
Qy 490 NNLKSPASDDLEIEIEKEGAFPMNRIETLEBKOKLVITIGDAAFHINHIYAIVLPSVQBIG 549
Db 481 NNLKSPASDDLEIEIEKEGAFPMNRIETLEBKOKLVITIGDAAFHINHIYAIVLPSVQBIG 540
Qy 550 RSAPFRQGANLIFMGSKVKTIGEMAFLSNRLEHLDLSEKQLTEIPVOAFSDNALKEVL 609
Db 541 RSAPFRQGANLIFMGSKVKTIGEMAFLSNRLEHLDLSEKQLTEIPVOAFSDNALKEVL 600
Qy 610 LPASLKTIREEAFKQNLKQLEVASALSHTAFNALDDNDGDEOFNKVVVTHNSYALA 669
Db 601 LPASLKTIREEAFKQNLKQLEVASALSHTAFNALDDNDGDEOFNKVVVTHNSYALA 660
Qy 670 DGEHFIVDPDKLSTSTVLDLEKILKEGLDYSTLRQTQTQFRDMTTAGKALLSKSNLRQ 729
Db 661 DGEHFIVDPDKLSTSTVLDLEKILKEGLDYSTLRQTQTQFRDMTTAGKALLSKSNLRQ 720
Qy 730 GEKQKFLQEAQFFLGRVLDLKAIAKAERKALVTKATNGQLLERSINKAVLAYNSAIKK 789
Db 721 GEKQKFLQEAQFFLGRVLDLKAIAKAERKALVTKATNGQLLERSINKAVLAYNSAIKK 780
Qy 790 ANVRLSEKELDLITGLVEGKGPLAQATMWQVYLLKTPPLPEYIYGLNVYFDSKGLIY 849
Db 781 ANVRLSEKELDLITGLVEGKGPLAQATMWQVYLLKTPPLPEYIYGLNVYFDSKGLIY 840
Qy 850 ALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLDIKTILNSKLSQLTSI 909
Db 841 ALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLDIKTILNSKLSQLTSI 900
Qy 910 RQVPTAAYHRAGIFQAIQNAABAEQLLPKPGTHSEKSSSESANSKORGLQSNPKTNRG 969
Db 901 RQVPTAAYHRAGIFQAIQNAABAEQLLPKPGTHSEKSSSESANSKORGLQSNPKTNRG 960
Qy 970 RSHAILPRTGSGKSFVYGILGYTSVALLSLITAIKKKY 1008

```
Db 961 RHAIPRTSGSGSFYGLVTSVALLSIAAIAKKKY 999
|||||
RESULT 5
QBE1D6 PRELIMINARY; PRT; 1055 AA.
ID QBE1D6
AC QBE1D6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cell wall surface anchor family protein.
GN OrderedLocusNames=SAG0421;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tetellin H., Masiagnani V., Cieglewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014211; AAM93327.1; -
DR TIGR; SAG0421; -
DR InterPro; IPR007093; LRR_Tp.
KW Complete proteome.
SQ SEQUENCE 1055 AA; 118129 MW; 2DD8B8715B1774E CRC64;

Query Match 74.3%; Score 3776.5; DB 2; Length 1055;
Best Local Similarity 71.7%; Pred. No. 5.9e-158;
Matches 756; Conservative 105; Mismatches 140; Indels 53; Gaps 5;

QY 2 KKHKTVALTLTVSVTHNQVPSLVKPEILKQTAQSSISGADYAESGSKSLKINET 61
|||
Db 3 KKHKTALALTTSVTVSYQEVYGLREESYKQEQTQSA-SEDDWPFEDNERKTNVSK 61
|||
QY 62 SGFVDDTVTDLPDSKRTTPEKIKONLAKGPRQELKAVTENT-ESEKQITSGSQLEQSK 120
|||
Db 62 NSTVDETVDLPFSDGNSNNSSKTESVSDPKQPKAKPEVTEASNSNSDASKVEVPKQ 121
|||
QY 121 SISLNTKVPSTSNWEICDPIKGTNTVLGSKGVEKLSQTDHLVLPQOAGDTQLIQVAS 180
|||
Db 122 DTASKKETLETSTWAKQDFTVRGDTLVGFSGSKGINKLSQTSHLVLPQOAGDTQLTQVAS 181
|||
QY 181 FAFTPDKKTATAEYTSRAGENCEISQDLDGKEIINEGEVFNLSYLLKVTPTGKYHIGQ 240
|||
Db 182 FAFTPDKKTATAEYTSRAGENCEISQDLDGKEIINEGEVFNLSYLLKVTPTGKYHIGQ 241
|||
QY 241 DAFVDMKNIAEVLNPLESLETISDYAPAHALKQIDLPNLKALGELAFDNDQITCKLSLP 300
|||
Db 242 DAFVDMKNIAEVLNPLESLETISDYAPAHALKQIDLPNLKALGELAFDNDQITCKLSLP 301
|||
QY 301 ROLMELAEAFKSNHIKTEPRGNSLVKIGEASFDNDLSQMLPDGLEKIESEFTGNP 360
|||
Db 302 RHLLKLAERAFKSNRIQTVEFLGSKLVIGEASFDNDLRNMLPDGLEKIESEFTGNP 361
|||
QY 361 GDHYNNRVLMVTKSGKPSGLATENTVYVNDKSLWQSPEDYTKWLEEDFTYQKNSVT 420
|||
Db 362 GDEHYNNQVLETRTGQPHQLATENTVYVNDKSLWQSPEDYTKWLEEDFTYQKNSVT 421
|||
QY 421 GFSNKGLOKVRNKNLEIPKQHGNTITEIGDAPFRNVDFQSKTLRKYDLBEIKLPSTIR 480
|||||
```

Best Local Similarity 72.0%; Pred. No. 9.8e-158;		Matches 756; Conservative 103; Mismatches 146; Indels 45; Gaps 5;	
Qy	2	KKHLKTVALT	TTVSVVTHNQVPSLVKEPILKQTOQASSISGADYAESGSKSLKINET 61
Db	3	KKHLKTLALAT	TVSVVTVSQEVYGLERBESVKQEQTQSA-SEDDWFEEDNERKTNVSKG 61
Qy	62	SGPVDDTV	DLPSDKRTTPEKIKONLAKGPREBELKAVENT-ESEKQITSGSQLEBQKE 120
Db	62	NSTVDET	VSDFSGNSNNSSKTESVSDPKQVPKAPPEVTQEASNSNDASKVEVPKQ 121
Qy	121	SLSLNTK	VPSTSNWEICDFITKNTLVGLSKSGVEKLSOTDHLVLPQAAADGTOLIOVAS 180
Db	122	DTASKET	ETIETSTWEAKDFVRGDTLVGFSKSGINKLSQTHLVLPQAAADGTQTQVAS 181
Qy	181	FAFTPDK	KTAIAEYTSRAGENGESIQDLDVKGKEIINEGEVFNLSYLLKKVTIPTGYKHIGQ 240
Db	182	FAFTPDK	KTAIAEYTSRAGENGKPSRLDIDQKEIIDEGEIFNAYQLTKLTPNGYKSIQ 241
Qy	241	DAFVNK	IAEAVNLPSLETISDYAFAPALAKQIDLPNKLKAIGELAFDNOITQKLSLP 300
Db	242	DAFVNK	IAEAVNLPSLETISDYAFAPALAKQVLPNKLKVGELAFDNOIGGKLYLP 301
Qy	301	ROLMLA	ERAPKSNHIKTEPRGNSLKVIGEASFODNDLSOLMLPDGLEKIESAFTGNP 360
Db	302	RHLIKLA	ERAFKSNRIQIVFELGSKLVIGEASFODNNLSNMLPDGLEKIESAFTGNP 361
Qy	361	GDDHNN	RVLWTKSGKPSGLATENTYVNDKSLWQESPEIDYTKMLEEDFTYQKNSVT 420
Db	362	GDEHNN	QVLRTRTGQNPQLATENTYVNDKSLWRATPDMDYTKMLEEDFTYQKNSVT 421
Qy	421	GFSNKG	LQKVRKNKLEIPKQHGVTIIEIGDNAPRVDPQNTKURKYDLEVKLPSTIR 480
Db	422	GFSNKG	LQKVRKNKLEIPKQHGVTIIEIGDNAPRVDPQNTKURKYDLEVKLPSTIR 481
Qy	481	KIGAFAP	OSNNLKSPEASDDEIEIEKGAFMNRIETLEKOKLVIGDAAFHINHIYAIV 540
Db	482	KIGAFAP	OSNNLKSPEASEDEIEIEKGAFMNRIETLEKOKLVIGDAAFHINHIYAIV 541
Qy	541	LPESVQ	EIGRSPFRONGANNLIFMGSVKVTLGEMAFSLNRLEHLDLSBQKQLEIPVOAF 600
Db	542	LPESVQ	EIGRSPFRONGALHLMFVGNKVTIGEMAFSLNKLESVNLSEKQKLTIEVQAF 601
Qy	601	SDNALKE	VLLPASLKTIREBAPKQHLKQLEVASALSHIAFNALDDNDGDSQFQPNKVVK 660
Db	602	SDNALSE	VVLPNLQTIREEAPKRNHLKEVKGSTLSQIAFNAFQNDGDKRFGKVVYR 661
Qy	661	THNSYAL	ADGHEFTVDPKLSTTVOLEKILKLTIEGLDYSTLRQTTOTQPRDMTTAGKA 720
Db	662	THNSHML	ADGERFIIDPKLSTTVOLEKLVKLTIEGLDYSTLRQTTOTQPREMTTAGKA 721
Qy	721	LLSKNLR	QGEKQKFLQBAQFPLGRVDLDKATAKAELAVTKATKNGQLLERSINKAVL 780
Db	722	LLSKNLR	QGEKQKFLQBAQFPLGRVDLDKATAKAELAVTKATKNGHLLERSINKAVL 781
Qy	781	AYNSAI	KANVRLKELDLTLGLVEGKPLAQATWQGVYLLKTPLPPEYIYGLNVY 840
Db	782	AYNSAI	KANVRLKELDLTLGLVEGKPLAQATWQGVYLLKTPLPPEYIYGLNVY 841
Qy	841	PKSGKL	IYALDMSDTTIGEGKDAYGNPLNVDEONEGVHALAVATLADYGLDITKTLN 900
Db	842	PKSGKL	IYALDMSDTTIGEGKDAYGNPLNVDEONEGVHALAVATLADYGLYIKDILN 901
Qy	901	SKLSQTS	SIRQVPTAAYHRAGIFQIAONAAAEAEQLPKP-----GTHSEKSSSESANS 955
Db	902	SSLDKI	KAIQIPLAKYHRLGFPQAIRNAAAEADRLIPKTPKGYLNTVPNYKQKQVEKNS 961
Qy	956	KD-----	-----RGLQSNPKTN-----RGRHSAILPR 977
Db	962	RPVDYK	TFIDKALPNEKVGDRVAKGHINIAETNNYVAVTPIRSEQQHLKSQSDVNLQP 1021
Qy	978	TGSGKSF	VYGLGYSVALLSLITAIKKKK 1007

Db	1022	TSSKNFI	EILGYVSLCLLFLVLTAGKKGK 1051
RESULT 7			
Qy	068831	PRELIMINARY;	PRT; 1081 AA.
Db	068831	PRELIMINARY;	PRT; 1081 AA.
AC	068831	01-AUG-1998	(TrEMBLrel. 07, Created)
DT	01-AUG-1998	(TrEMBLrel. 07, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	Surface antigen BspA.		
GN	Name=BspA;		
OS	Bacteroides forsythus.		
OC	Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;		
OC	Porphyromonadaceae; Tannerella.		
OX	NCBI_TaxID=28112;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC43037;		
RX	MEDLINE=99043895; PubMed=9826345;		
RA	Sharma A., Sojar H.T., Glurich I., Honma K., Kuramitsu H.K.,		
RA	Genco R.J.;		
RT	"Cloning, expression, and sequencing of a cell surface antigen		
RT	containing a leucine-rich repeat motif from Bacteroides forsythus ATCC		
RL	43037.";		
RL	Infect. Immun. 66:5703-5710(1998).		
DR	EMBL; AF054892; AAC82625.1; -.		
DR	PIR; T31094; T31094.		
DR	InterPro; IPR003343; Big_2.		
DR	InterPro; IPR007093; LRR_Tp.		
DR	Pfam; PF02368; Big_2; 1.		
DR	SMART; SMO0635; BID_2; 1.		
SQ	SEQUENCE 1081 AA; 113907 MW; 65E36FA0ACBD5C CRC64;		
Query Match 6.5%; Score 331.5; DB 2; Length 1081;			
Best Local Similarity 25.3%; Pred. No. 2.4e-06;			
Matches 147; Conservative 83; Mismatches 206; Indels 144; Gaps 23;			
Qy	110	TSGQLSQSKESLSLN-----KTVFSTSNWEICDFITKNTLVGLSKSGVEKLSQTDH 162	
Db	19	TLGATAGNSGTTGLNWSYSGTGTALITGTGAMPDF---NNASEIPMHSLSQSKIQT-- 72	
Qy	163	LVLPSQAADGTQLIQVASFAT-----PDKKTAAEYTSRAGENGESIQDLDVQKE 213	
Db	73	-----VTIGDVTSVGNNAFSDCALTSVTLPSNLTAIGDHAFK-GCSGLTSITIPNSVT 125	
Qy	214	IINEGEVFNLSYLLKKVTIPTGYKHIGQDAFVDNKNIAEAVNLPSLETISDYAFAPALA-LK 272	
Db	126	TIGEWAPKGCGLKSLITLPSNLTAIGQSALSGCTGLTSITIPNSVTTIGEWAFGCGLT 185	
Qy	273	QIDLPDLNKAIGELAFDNOITGKLSLPQMLRLAERAPKS-NHIKTIETFGNSLKVIGE 331	
Db	186	SITPNSLTAIGESAFYCGGALTSITLPDALTTIGESAFKGCGLKSIITF-PNSLTITGE 244	
Qy	332	ASPOD-NDLSQMLPDGLEKIESAFTGNPDHNNRVLVLTSGKSNPSGLATENTYVN 390	
Db	245	SAFYDCGALTSITLPDALTTIGRSAPFYG-----CSGL----- 276	
Qy	391	PKSLWQESPEIDYTKMLEEDFTYQKNSVTGFSNKGKLVKRNKNLEIPKQHGVTITEI 450	
Db	277	-----KSITFPN-----SLTTI 288	
Qy	451	GDNAFRNVDFQNTKRYDLEEVKLPSTIRKIGAPAFOS-NNLKSFASDLEIKGAP 509	
Db	289	GESAFYNGC-----SLTSITIPNSVTTIGRSAPFYCGGLKSIITLPDGLTTIERAF 339	
Qy	510	MNNRIET-LELKOKLVITGDAAFH-INHIYAIVLPESVQEIGRSPFRONGANNLIFMGSK 567	
Db	340	YNCQGLTSITIPNSVATIGESAFYCGGLKSIITLPDGLTTIEWGAFYNGCGLTITIPNS 399	
Qy	568	VKTLGEMAFSL-NRLEHLDLSEKQLEIPVQ--AFSDNALKEVLL--PASLKTIREBAP 622	
Db	400	VSTIGESAFYCGGALKQDVTVAWD---TPIDIQDVFRRLTSGIRLHVPAKKTIVYE--- 453	


```

QY 623 KQHHLKOLEVASALSIAFNALDDND-GDEQFDNKVVVKT 661
DB 454 AKDVKWE-----FNIVEDDDFGGLQWNYDAATKT 482

RESULT 8
Q8TI64
ID Q8TI64 PRELIMINARY; PRT; 1995 AA.
AC Q8TI64;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Cell surface protein.
GN OrderedLocusNames=MA4292;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galleon J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McSwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011143; AA007636.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR010983; EF_Hand_like.
DR InterPro; IPR007093; LRR_Tp.
DR InterPro; IPR00601; PKD.
DR SMART; PF00801; PKD; 6.
DR SMART; SM00089; PKD; 6.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS50093; PKD; 6.
KW Complete proteome.
SQ SEQUENCE 1995 AA; 208421 MW; 031C7F16092BE035 CRC64;

Query Match
Best Local Similarity 6.2%; Score 317; DB 2; Length 1995;
Matches 177; Conservative 112; Mismatches 289; Indels 190; Gaps 36;

QY 14 TVSVVTHNQEVFLVKEPKILKQFQASSISGADYA-----ESSGSKLKI 58
DB 466 TVSLTVSN-----TGGSSSTETKADYIVAALPDSPVADSVDTVSGSAPLTV 511
QY 59 N-----ETSGPV-----DDVTDLFSKRTTPEKIKDLNLAGPREQELKATENTESK 107
DB 512 QFTDLTNGFVSWANDFSDGIDSYEQNPWSYTNAGNYVTWTATNLN--VASDPTVTRV 569
QY 108 QITSGSOLQEQ-----SKESLSLNTKVPSTSNWEICD----- 138
DB 570 NLTSVENLPMADFSADPVATLIGCSVQFTDLSNSTPSQWQDFNNDGTIDTMQNPSYT 629
QY 139 FITKGNLTVLGLSKS-----GYEKLSTQDHLVLPQQAADGTOLIQVASFAFTPD-KKTAIAFY 194
DB 630 YTTAGTYTVNLTVSNPAGSGDEKTDYIVVKEQAS-----HASDFTYSDGSSITITEY 683
QY 195 TSPAGENG-EISQLDVGKEI--INEGEVFNYSLLKKVTIPTGYKHIGQDAFVKNKNAE 251
DB 684 T---GSDGVIIPAEIEGLPVTITIGASAFYGCASPTTVPITPNSVTTIGDSAFQDCSALT 740

```



```

Db 650 EGCTALTSITIPDSVTSIDRYAFRGCTALTSATIGSGVTGIGVGFYQCTNMASVTIPD- 708
Qy 394 SLWQESPEI--DYTKWLEDDFYQK--NSVTGFSNKGKLVKRNKNLEIPKQHNGVTITE 449
Db 709 ----NTEIGNDAFWHC--TNLISMTIGNVTTIGSFAFGCTGLASVTIPD-----SVTT 758
Qy 450 IGNAFRNVDYFONKTLRKDYDLBEVPLPSTIRKIGAFQFOS--NNLKSFEASDDLEBEIKEGA 508
Db 759 INDPAFSGCT-----NLTSITLPNSVISIGNFVDFQDCSALTSVTFSGSLTSGSNV 809
Qy 509 FNN-----NRIETLELKDVLVTTGDAAFHINHIAIVL-----PESVOETGRSAP 553
Db 810 FOSCTALNAI--NVDANNVTYTSIDGIVYKNDVTAVVLPFSKAGSIVIPDSVTSISSYAF 868
Qy 554 RQNGANNLIPMGSKVKVTIGEMAFPLSNRLEHL--DLSEQKQLTEIPVQAFSD-----602
Db 869 YGCAGLTSVTIGNSVTITGQAF--NGCTALTSITIPDRVTEIGICAFQDCSALTSATFG 926
Qy 603 -----NA-----LKEVLLPASLKTIREBAFKK--NHLKOLEVASALSIAFNALDDN 647
Db 927 TGLITIGENAFNGCTNLTSAMIPNNVTITIGKSAFNGCTNLTSVMIGSGVTSIGANAKIAP 986
Qy 648 DGEQPDNKKVVKTHNSVALADGHEHFIVDPDKLSSTIV 686
Db 987 HGCTAL--TEVSVDAINTAYSSIDG-----VYDKAGTTLV 1020

RESULT 11
Q7PG02
ID Q7PG02 PRELIMINARY; PRT; 1066 AA.
AC Q7PG02;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000023103 (Fragment).
GN Name=ENSANG00000019883;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxId=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAA01008846; EAA45120.1; -.
FT NON_TER 1 1066
FT NON_TER 1066 1066
SQ SEQUENCE 1066 AA; 122246 MW; 9DE10C52881P98BA CRC64;

Query Match 4.9%; Score 247; DB 2; Length 1066;
Best Local Similarity 22.1%; Pred. No. 0.012;
Matches 205; Conservative 147; Mismatches 350; Indels 226; Gaps 41;

Qy 55 KUKINETSFPDVTVDLFSKRTTPPEIKDKNLAKPREQELKAVTENTSEKQITSGS- 113
Db 242 KTSMBESNAGLQSLDALEHTTKQLADNLEQEMAK--TADNRLTEELEAEKKQLMQEL 298
Qy 114 --GLEQSKESLSLUNKTVPSTSNWEICDPITKGNLTVGLSKSGVEKLSQTDHLVLSQAAD 171
Db 299 ETQKKQLMQELEVAKTSMSNNAEL-----QRSKAEIQT-----KMQLAD 339
Qy 172 GTQLIQVAFAPFPDKKTAIAEYTSRAGEISQLDVGKTEINGEVFNSVLLKKVTI 231
Db 340 NLE-----KEIAKTADNLRRTTEELEVEKKQLTQLEVAKT-----374
Qy 232 PTGYKHIGODAFVNDKNIAEAVNLPESLETISDYAFHAHLAKQIDLPDNL--KAIGELAFPD 290
Db 375 -----SKSENIAE--LQRSLEAIEQ-----TKQLADNLEKSEIAKTADL- 411

```

```

Qy 291 NOITGKLSL--PROLMRLAERAFKSNHKTIBFRGNSLKVIGEASFQDNDLSQMLPDGLE 349
Db 412 NRTEKLEVEKKQLMQELEVA--KTSMSENIAELQRSLEAI-----EQTKLOPADNLE 462
Qy 350 KIESAFTGNPGDDHNNRVVLTWTKSGKNPGLATENTYVNPDKSLWGESPEIDYTKWLE 409
Db 463 KEIAKTADNLRRTTEELEVEKKQLTQLEBAKSSGSENI-----AELQSLKAIEQTKLQL 517
Qy 410 EDFTYQKNSVTGFSNKGKLVKRNK-----NLEIPKQHNGVTITE-----I 450
Db 518 ADNLEKEIAKTADNLRRTTEKLEVEKKQLMQELEVAKTSMSENIAELQSLKAIEQTKLQL 577
Qy 451 GDNAPRNV-----DFONKTLRKDYDLBEVPLPSTIRKIGAFQFOSNN-----LKSF 495
Db 578 ADNLEKEIAKTADNLRRT--EKLEVEKKQLMQEL--EVAKTSMSENNAELQRSKALELTKL 635
Qy 496 BASDDLE--EIKGAPMNNRIETLEKDKLVITGDAAFHINHIAIVLPESVOEIGRS--A 552
Db 636 QLADNLEKEIAKTADNLRRTTEELEVEKKQL-----LQELEVAKTSMSENIAELQSLKA 689
Qy 553 PRONG-----ANNLI PMGSKVKVTIGEMAFPLSNRLEHLDLSEQKOLT--EIPVQAFS-----601
Db 690 TEQTKLQLADNL-----EKSIKTAELNRRTEELEVE--EKKQLTQLEVAKSSSENIAT 742
Qy 602 -----DNALKEVLLPASL--KTIREBAFKKKNHLKOLEVASALSIAFNAL 643
Db 743 LQRSLEAIEQTKLQLADNLEKEIAKTADNLRRTTEELEVEKKQLQLEVAKTSKSEIAE 802
Qy 644 LDDN--DGDEQFD-----NKVVVKTTHNSVALADG-----HFIVDPDKLSSTIVD 687
Db 803 LQOSLEAIEQTKLQLADNLEKEIAKTADNLRRTTEELEAEKKQLMQELEVAKTSMSENNAE 862
Qy 688 LEKILKILLEGDLYSTLRQTTQTFQDMWTAGKALLSKSNLR-----QGEKQKFLQEAQFPL 743
Db 863 LQRSLEAIE-----QTKQLADNLEKEIAKTAEINLRRTTEELEVEKKQLTQBLE--V 911
Qy 744 GRVLDLDAKIAEAKAEKAVTKKATKNGQL-----LERSINKAVLAYNNSAIKKNVRLKEKELD 800
Db 912 AKSSGSENIABEQSLKAIEQTKL--QLADNLEKEITKTADNLRRTTEELEVEKKQLTQELD 970
Qy 801 LLTGLVEGKGPLAQATMVQGVYLLKTPPLPEYIYGLNVYFDKSGKL-----IYALDMSPTI 857
Db 971 DM-----KTTNLSRQSLVLE--LTKTQITNSLELVLKIGDLQHLKAMDEAKTL 1020
Qy 858 GEGQKDAYGNPIL-----NVDEEDNE 877
Db 1021 AEKLGDSERSEMLAKISLLQQLNEERE 1048

RESULT 12
Q8TI71
ID Q8TI71 PRELIMINARY; PRT; 1743 AA.
AC Q8TI71;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell surface protein.
GN OrderedLocusNames=MA4285;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxId=22114;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G. MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Ainoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McSwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grzame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,

```


QY 289 -FDNQITGKSLPQRLMLRAERAFKSNHIIKTIEFRGNSLKVIGEASFQNDLSQMLPDPG 347
Db AIDNAVAGKLEINDSLTNEEQAYVDLIN-----NEADNAQKIAEATTPPE 4080
QY 348 LEKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4081 VTRAQEE-----GVKNIIININVTPTSPAKDAANAII-----DQALNKKKDEINNAATN 4127
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 408 L-----EEDFTYQKNSVTGFSNKGLOKVKKNKLEI-----PKOHNGVTITEIGDAFRNV 458
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4128 ISSEKTDLIKATEAANTAKNINNATTNNEVETAQVDGEKAIADVTVPGLSDIKKSI 4187
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 459 DFQNTKL-RKYDL-----EVLKSTIRKIGAFAPQSNLKSFEASDLBEIKEG 507
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4188 DLINKALNEKQDEINNASLSQDEKQELIDQAKIATEAI-----NEINNAQTNDKEAADT 4245
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 508 AFWNNRIETLEKDKLVITIGDAAPHINHIYAIVLPESVOEIGRSFQNGANNILFMSGK 567
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4246 GVKN-----IENVSP-----SIEDAKONATQAIIDALNSKKNEI-----NNASNL----- 4286
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 568 VKTLAGEMFLSNRLEHLDLSEKQKLTETPQAFSDNALKEVLLPASLKTIREEAFKKNHL 627
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4287 ---TDSEKTDLIN-----QATEIA-----NAKDAINSATTNTAVEAAYKG--- 4325
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 628 KOLEVASALSHIAFNALDNDGDEQPDNKKVVKTHNSVALADGEHFIYDPDKLSTIVD 687
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4326 ----VADINNIHFTNLDS-----KKAANS-AIEDALNTKKDEINNASLSD 4367
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 688 LEKILKLEGLDYSLRTQTOQFRDMTTAGKALLSKSLRQ-----GEKQKFOE----- 738
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4368 SEK-AKLI-----NQATEIANAARAANNATTNSAVTAAENKGIEDIANINV 4413
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 739 ---AOFFFLGRVLDKAIKAERKALVTK---KATNGOLLERSINKAVLAYNSAIIKAN 791
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4414 PSIAETKQAIIDAIQOVQAKNSQTEAKNLSADEQKLNIDQ-VNKIA-----QDAINKLN 4468
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 792 -----VKRLEKELDLTL-----VEGKGPQAQATVGVVLLKLTPLPL 830
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4469 DPATTNTEVITTRDKAIDQITNLFIPTLDSVQKQAQEAQKIDEI----- 4519
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 831 PRYIYGLVYFKSGKLI-----YALDMSDTTGEQOKDAYGNPILND-----EDNEGY 879
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4520 -----NKADNLTDQMKQNLIDQVDQVADKATKAINNAQTNDVKEAEIEGLEDI 4568
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 880 HALAVATLADYGLDITKT-----LNSK-----LSQTSIRQVPTAAVHR 919
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4569 DSIKVPSLVEKODAIKENDALKKKTDEINADLDQKQKDELIQITDI-----ATETK 4623
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 920 AGIFOAIQNAAEAE-----QLLPKPGTHSEKSSSESANSKDRGLQSN---PKTNR 968
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4624 TKVFNATTNAEVDAAEAGIKAEAVKI PARTADNSNT--ESESKEQVITVNSVQPKRNA 4681
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 969 GRH-----SAILPRTGSK-----GSFVYGLGVTSVALLSLTAIKKKK 1007
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407
QY 4682 VHHKNGTPVKNKATLPQTKGKNSNLTLAGAALLGLAGVFS--LFLGLGDKRKNK 4734
Db LKISEAETGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407

RESULT 14

Q81J57 PRELIMINARY; PRT; 1139 AA.
AC Q81J57;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE QF122 antigen.
GN ORFNames=PF10_0115;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,

RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteua M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Vaidya A.B.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.,
RA "Genome sequence of the human malaria parasite Plasmodium
falciparum".
RT Nature 419:498-511 (2002).
RL Nature 419:498-511 (2002).
CC -!- SIMILARITY: Contains 2 KH domains.
EMBL; AEO14830; AAN35313.1; -.
DR HSP; Q9UNW9; 1DTJ.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH 1; 3.
DR SMART; SM00322; KH; 4.
DR PROSITE; PS00084; KH TYPE 1; 2.
SQ SEQUENCE 1139 AA; 131625 MW; 9983227714206714 CRC64;

Query Match 4.5%; Score 230.5; DB 2; Length 1139;
Best Local Similarity 18.9%; Pred. No. 0.073;
Matches 219; Conservative 173; Mismatches 422; Indels 347; Gaps 45;
QY 44 GADYAESGSKGKIKINETSQVDDTVTDLFSDKRTTPEKIKONLAKGPREQSL----- 96
Db 2 GRKATTSSILEKPKVKNKSGVNNMKDMDKQGGSKL--NEKKGSKVTSLTNAHNN 59
QY 97 ---KAVTEN--TESEKQITSGSL-----OSKESLSLNTKTPVSTSNMEI 136
Db 60 GSIQKKEELENKLNHNKLEENCVKSEKVEVKGQDQKNEQAKKNNNNNN--- 116
QY 137 CDTFTGNTLVGLSKSGVEKLSQTHLVLPQADGTQLIQVASFATPDKKTAEYTS 196
Db 117 -----NKKGTGTENDMDKINQSGTNDKDKKANNTTE-----GNEKNKSGNNKKYVG 165
QY 197 RAGENGESQLDVDCKEI-----INEGEVFNLSYLLKKVTIPTGYKHIGODAFVDKN--- 248
Db 166 NKDENMKVELMDVTYKNNSTGTGINSSNNNNNNKNNNNNNNNNNNNNNNNNNNN 225
QY 249 -----IAEVNLPESLETISDYAFAPHLAKQIDLPDLNKAIGELAPF 289
Db 226 IISNNNNNNNSDVKKDKKEIKLEDERKSENDLKKFKV-LNKADNYKINDK--- 282
QY 290 DNQITGKLSLQRLMLRAERAFKSNHIIKTIEFRGNSLKVIGEASFQNDLSQML 344
Db 283 EAEIKKKEI-----ENFISTINSIRFNGQTTPT-----NMKLSADM 323
QY 345 PDGLEKIEEFTGNGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDY 404
Db 334 ENKCKDAKSKKKKINP-----KNLEETEEANTYIN----- 353
QY 405 TKWLEEDFTYQKN--SVTGFSGNGL-----QKVKN-----KNLE 437
Db 354 --YLEEQSLITKYNENFKNFQRLISLRKICEELKENLKNLKNLKNLKNLKNLKNL 411
QY 438 IPKHNGVTITEIG-----DNA-----FRNVDQNTLRKY-----DLBEVKLPSTIR 480
Db 412 MKKEKQNVDTIETADIIIRKDYAIPDTYTNLMKPNFLNRIQTKYFVYVNDNKLSTNNN 471
QY 481 KIGAFAPQSN-----LKSFE-----ASDLE-----EIKEGAFM 510
Db 472 MNN 531
QY 511 NNRIETLEKDKLVITIGDAAPHINHIYAIVLPESVOEIGRSFQNGANNILFMSGK--- 567
Db 532 NKKIFKI-----MLNWDGSP-----KQWEDTVVVFVHDVNDMLYCYCKKDDI 574
QY 568 ---VKTLAGEMFLSNR--LEHLDLSEKQKLTETPQAFSDNALKEVLLPASLKTIREEAFK 624
Db 575 ENLKELIKATNNEKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKN 628

```

QY 625 NHLKQLEVASALSHIAFNALDNDGDEQ-----FDNKVVVKTHN-SYALA 669
DB 629 NYDATFASATIKSGKNS-DIEAEAEKLEILKNLDSFVPFBEKEIFSLYKCKSYELN 687
QY 670 DGEH----FIVDPDKSLSTIVDLEKILKILIEGLDYSTLRQTQTOTQFRDMTTAGKALLSKS 725
DB 688 EIRKVLNFIPIRODNGLSVGKKONITALEILEYSKNIISNKTVKKLTTEE-EAFLPNS 746
QY 726 NLRQSEKQKFLQEAQFFL-----GRVDDLKAIKAEKALVTYKATKNGQLLERS 774
DB 747 NYRQGIKAKTGAEVKIFKNTNHKELNLSGNKNDIDKALEMIEELLKKEKCT-----QVD 800
QY 775 INKAVLAYNNNAIKKANVKRLEKEL-----DLTLGLVEG 808
DB 801 INEKVIALLSL--KAQIKIEKDEKCTCTSIQIKNTSHVAIYGHEDNIHLAKDVLNLSVQS 859
QY 809 KGPLAQATWVGQVYLLKTPPLPEYI-IGLVYFDKSKLIYALDMSDTIGEGOKDAYGN 867
DB 860 EGKEGK----EGKF-----IPNNLYVTVMVETEHISGVIG--KKGRTINKIQEDTFAK 908
QY 868 PILNVDEDEGHALAVATLADYEGLDIKTIILNSKLSQTSIRQVPTA-----AYH 918
DB 909 KI-HIDKENKKIYIHGTQKTVDAQKEIQILNRSKEEMNNNNNNSSNGGAFNNSYH 967
QY 919 RAGIFQATQNAAEABEQLLPFGTHSEKSSSESANSKORGL-----960
DB 968 ESRVNNRSHNSSSAHHRKSHKTSRSRPSRSHKSESCKGYIINTNDEKAPPSLHDVTN 1027
QY 961 -----QSNPKTNRGRH 971
DB 1028 MQSKKKNKLSQSNQTTNTKKH 1048

RESULT 15
Q8EWP8 PRELIMINARY; PRT; 3317 AA.
AC Q8EWP8,
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE Predicted cytoskeletal protein.
OS OrderedLocustNames-MVPEI550;
OS Mycoplasma penetrans.
OC Bacteria; Filumicutes; Mollicutes; Mycoplasmatataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RC STRAIN=HF-2.
RX MEDLINE=2354719; PubMed=1246555; DOI=10.1093/nar/gkf667;
RA Sasaki Y., Iihikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Morino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL; AP004170; BAC43946.1; -.
KW Complete proteome.
SQ SEQUENCE 3317 AA; 385843 MW; 0641151BAPA992BB CRC64;

Query Match 4.5%; Score 229; DB 2; Length 3317;
Best Local Similarity 20.1%; Pred No. 0.38;
Matches 222; Conservative 199; Mismatches 406; Indels 280; Gaps 53;

QY 57 KINETSQGVDDTVTLDFSDKRTTPBKIDNLAQPRE-----93
DB 1782 ELNNIISNIDSEFKLISFSDSTNQIGENLVKDIKHSNKHAIKWDSSIMEMNSKFDRL 1841
QY 94 -QELKAVTENTSEKQITSGSQLESLSLNTKTPVTSNWEICDPITKGNLTGLSKS 152
DB 1842 IEEMKVQTKIENSNSVNDTNLMIESKNQNE-----EIPNYVSL-----LEKS 1884
QY 153 GVEKLSQTDHLVLPQAADGTQLIQVAFPTPK-KTAIAEYTSRAG-ENGEISQLDVD 210

```

```

DB 1885 -----TSSLVNEMOQA-----QL-----DKEKIKLKELETLIDLQNSIESLIDD 1924
QY 211 GKELINE-----GEVFNSSYLLKKVTIPTGY-----KHIGQDAFVDMKNIAEVNLP 256
DB 1925 KNDPIEBEINFLNOKELFNKENFVLENSTIGNIISAFDKKITDLEFNFORKLSNLDL-S 1983
QY 257 SLETTISYAFALHAKQIDLPDLNKAIGELAFPDNQITQKLSLPRQMLAERAFKPNHI 316
DB 1984 SLKNLFDODKOSTLASVINSVEDNLSNVEI-----LRNELK--ELNKILEDISEANSI 2035
QY 317 KTI-----EFRGNSLKVIGEASFQNDLSQLMPLDGLKIESE--AFT 357
DB 2036 NEINRNSOPKLEBVSQPDQFR-DVLKTYVEDIIHKSENNQIIVDDLEKINDEFPSLI 2094
QY 358 GNPQDDHNNRVVLWTKSGKNPSGLAT--ENTYVNPDKSLWQESPEID-YTKWLEE----410
DB 2095 QNLRSENELSNLMDKFNLSGLVERFNSPNINNKPDFIAKLDEKTKSPENRMD 2154
QY 411 --DFTYQKN-----SVTGFNSKGLQKVRNKLKLEIPKHNGVTTITEIGDNAFR---NVDF 460
DB 2155 LINFVYELNKEYQKNISSY-----IQKIESNN--EFISDLDAITLTKENDKLEYLELIDI 2209
QY 461 QNKTL-----RKVDLEEVKLPTIRKIGAFAPQSNNLKSFEA-----SDDLEE-IKE- 506
DB 2210 QNEIIVRLIDENKY-IDEVEYPLTNKK--EALDNKKLNDIESINNINNPEKIKCL 2265
QY 507 GAFNNRIETLEK-----DKLVITGDAAPHINHIYAVLPESVQIEGRSAFQNG 557
DB 2266 GENFEKVDSTELQNLGNLSESDK-KTLDVIRSQLSKNYELLSEE-----FKNNF 2315
QY 558 ANNLI PMGSKVKTIGEM--AFLSNRLEHLDLSEQKQLTEI-----PVQAFSDNALK 606
DB 2316 A-----SLIKSIENLYSSYGKDSLSQIASSELKFCYNEIGISLGISLNKQNFDFLIK 2368
QY 607 EV-----LLPASLKTIREEAFKKNHLKQLEVASALSIAFNF--ALDNDGDEQ-----F 653
DB 2369 ELETNKLILASQVDAMNLILOSNKINSKEIVSYENLKGNYANLSQKEKEQALIFKIF 2428
QY 654 D--NKVVVKTTHNSYALADGEHFIVDPDKLSSTIVDLEKILKILIEGLDYSTLRQTOTQF 711
DB 2429 DELNEIINIQNQDIGSLLEBRDYIIDBVQKPTNKIEBENSKNLIQSADFKNLKDYLKTSV 2488
QY 712 RDMTAGKALLSKNLQGEKQKPLQEAQFPLGRVDLDKATAKAEKAL-VTKKATKNGQL 770
DB 2489 DNLGNKIEGL--KSDL-----SNKFTSISOHL-----IDDSAGYENILKVEQNUTENYSP 2538
QY 771 LER---SINKAVLAYNNNAIKKANVKRLEKELDLTLGLVEGKGPLAQATWVGQVYLLKTP 827
DB 2539 LKKDLYSIFENI FNI IDELIKPSDVASFQKIDDLKEEDINNVNPILASV-----2587
QY 828 LPLPEYIIGLVNVPDKSGKLIYALDMSDTIGEGOKDAYGNP--ILNVDE-----DN-----876
DB 2588 -----DKLNKKNQIILDEIQAVITSGKCEANSNDQNYLKVEESVKINDNLELK 2633
QY 877 -----BGYHALAVATLADYEGLDIKTIILNSKLSQTSIRQVPTAAYHRAGIFQAIQNA 930
DB 2634 NKISKLDNWKDLVASVENKNGV--ISSFVNKLEBEISKINEVTPAPFDINAYNLIISKI 2692
QY 931 ABAEQLLPKPQTHSEKS-----SSSEGSANSKDRGLQSNPKTNRG--RHSAILPRTGS 980
DB 2693 SEISNLL-----NLSKKNLTGILNTVNARLSSENSDKVLQALKWFKNTILFINKTIEE---2745
QY 981 KGSFVYGILGYTSVALLSLITAIKKKK 1007
DB 2746 -----LNVENKYLELINANKNE 2764

RESULT 16
Q7RS28 PRELIMINARY; PRT; 1523 AA.
AC Q7RS28
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)

```

01-MAR-2004 (Tremblrel. 26, Last annotation update)
DT Hypothetical protein.
DE Name=PY00538;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Anguolli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii";
RL Nature 419:512-519 (2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC preliminary data.
CC EMBL; AABL01000149; EAA16762.1; --
DR InterPro; IPR003123; VPS9.
DR Pfam; PF02204; VPS9; 1.
KW Hypothetical protein.
SQ SEQUENCE 1523 AA; 177622 MW; 7F8FFA7218E933B6 CRC64;

Query Match 4.5%; Score 227; DB 2; Length 1523;
Best Local Similarity 19.3%; Pred. No. 0.16;
Matches 202; Conservative 150; Mismatches 410; Indels 284; Gaps 46;

QY 16 SVVTHNQEVFLVKE-----PILKQTOASSISGADYAESGSKLKINETSQVFD-- 66
Db 111 SLWESNESYIVKKNKERNKTPIKKQVNSDKLISS-----MNEISNPFVDQ 158
QY 67 -----DTVTDL-----PSDKRTPEKIKNL---AKGPREQE-- 95
Db 159 SNGQKLPDSTPDLWEKTEYDKBYELTGNESFYKKKKKKLDYIIHSGKGLNNSNI 218
QY 96 -LKAVTESKQITSGSLES-----KESLSLNK----- 126
Db 219 YKGGTENGNNVNGENGNNRIRVMSKSNRVKGDKEIKNSIDYHKCKSQNTL 278
QY 127 -TVPSTSNWEICDFITKGNLVLGSKSGVEKLSQTDHLVLPQQAADGTQLIQVAFPTP 185
Db 279 TKIDSFYBEEFTSLKNGKL-----NDDINLISEHD-IQIECYDKSISQCKVNNKISI 333
QY 186 DKYTAI-----AEYTS-----RAGENGISQLDVGDKEIINEGEVFNLYLKVTI 231
Db 334 KKKQDILLGDEYKNIYKKNKIDIFINSNDNNNNILSDSKHSISNLQINDKRAIKKTES 393
QY 232 PTG-----YKHIGQAFVNDKNIAEVLNPSL-----ETISDYAFALKALQIDLPD 278
Db 394 PNSTYSRYRIYIGNDNGNDNDLEKTKBKIPKKNISTQELINSYK-----DCTN 444
QY 279 NLKALGELAFFNQITGKLSLPQMLRAERAFKSNHKTIEFRGNS---LKVI--GBAS 333
Db 445 NSVDI-----LVDEQSKDEIQAKREIK-----KKKIMKPINNKTSQSFTSVIDSENT 494
QY 334 FQNDLSQLMLPDGLEKIE--SEAP-TGNPGDDHNNRVNVLWTKSG-----KNPSG 381
Db 495 YYDNTKFSQSSNLLKLDSESEFFCNSNMMDDEKKKIK--TKGTDKDKNLKXKKMED 552
QY 382 LATENTYNPKSLWQESPEIDYTKWLEDFTYQKNSVTGPNKGLQVK---RNKL-E 437
Db 553 MDNKKKEIEMEIEIKNKSEISQAEQSYFESIKTNNNNSKSFDEIKYLNKNKTE 612
QY 438 IPKQHGVTITBIGNAFNRVDFQNTLRKYDLEEVKLPSTIRKIGAFQSNLKSFEA 497

613 TNYNYSI-----DDINENFKLNKTL-----EIKKQTIK-----TNSKNTIC 651
QY 498 SDDLEETKEGAFMNNRIETLELKOKLVITGDAA-----PHINHIYAIVLPESVQETGRSAF 553
Db 652 SSDKEKIRNNSKEKJUVKTKERIKVIGDIADELEIEISNETYMKRENVECE--KKGTI 710
QY 554 RQNGANNLIFMGSK-----VKTLEMAFL-----SNRLEHLDLSEKQLTEIPVQAFSDNAL 605
Db 711 KQKGNENIKTNRDKKITENVKKISSEKPYDNSNSTWNSNFSNYSNLAKKKKKKISDK-- 768
QY 606 KEVLLPASKTIREBAFKQNHLEKQEVASALSHIAFNALDNDGDQEQDFNKVVKVTHNS 665
Db 769 -----ITVLHSTPNLSRINTNLIEBNKKNKDCDEKEI----- 800
QY 666 YALADGEHFIVDPKLSSTIVDEKILKLEGLDYSTLRQTQTQTFQDMTTAGKALLSKS 725
Db 801 YA-----DILVD-DK-----IIOKENIK-----KNDIKNVYIRDIQDGTANSNKF 841
QY 726 NLRQEKQKQLEQAQFFLGRVLDLQAKIAKAKALVTKKATKNGQLLERSINKAVLAYNNS 785
Db 842 IVK--KKNDFSIDGNETCGQEVDEVEKKEK--KEILENGQNVQNVQNVQNESF 897
QY 786 AIKKAN-----VKRLEKELDLTLGLVKGKPLAQATWV-----QGVYLK 825
Db 898 VIQGRKPKMFIKRESYKKBQVNEQVDENKIITKKEFIKWNEMNLINYNENKKEK 957
QY 826 TPLPLPEYI---GLNVYFD--KSGKLIYALDMSDTTIGEGKADKGNPILNVDENEGVH 880
Db 958 TESNLRFEIKKSTNNELNINSKVI-----EKYKNGLNN--EDNKIFN 1001
QY 881 ALAVATLADYEGLDIKITILNSKLSQ 906
Db 1002 NDTISENTSNNKKNATYNNFLES 1027

RESULT 17
Q8GLH0 PRELIMINARY; PRT; 178 AA.
AC Q8GLH0;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Immunoreactive protein Sello (fragment).
OS Streptococcus equi.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=11336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CF32;
RA Qin A., Artiushin S., Timoney J.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY137519; AAN18285.1; --
FT NON TER 178 178
SQ SEQUENCE 178 AA; 19194 MW; 20506DF6D5608BEA CRC64;

Query Match 4.4%; Score 225; DB 2; Length 178;
Best Local Similarity 35.7%; Pred. No. 0.0096;
Matches 66; Conservative 35; Mismatches 72; Indels 12; Gaps 6;

QY 1 MKKHKTVALTLTTVSVVTHNQEVFLVK-EPILKQTOASSISGADYAESGSKSKLIN 59
Db 3 IKKCPQILALFLAALASQQTWHLERQTSVKQAHESDDW-----FDEGEVSDIKSN 58
QY 60 ETSQVDDVTVDLFS--KRTPEKIKDNLAGPREQELKAVTENSEKQITSGSOLEQ 117
Db 59 -TPTEIDSTVTLEFQDAEPKTKAEK-BESKAPSPQEAALPA---ESQKEKTSKAKTDHA 113
QY 118 SKESLSLNKTVFSTSNWEICDFITKGNLVLGSKSGVEKLSQTDHLVLPQQAADGTQLIQ 177
Db 114 TVQAEAQEQSGSDSPWLADDFTVKGDTLVGLSKAGLAKLSKTPVLVLPFRIGDKTVLTR 173

```

Qy 178 VASFA 182
Db 174 IASFA 178

RESULT 18
Q7RPUO PRELIMINARY; PRT; 2719 AA.
AC Q7RPUO
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Rhostry protein.
GN Name=PY01365;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=1236865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiolini S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519 (2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000359; EAA20682.1; -.
DR HSBP; P03069; IGCN.
DR InterPro; IPR006499; ReticulocyteBP.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
SQ SEQUENCE 2719 AA; 318990 MW; AC3B69CEF74BB6E2 CRC64;

Query Match 4.4%; Score 224.5; DB 2; Length 2719;
Best Local Similarity 18.6%; Pred. No. 0.45;
Matches 207; Conservative 184; Mismatches 370; Indels 353; Gaps 52;

Qy 6 KTVALTITV-----SVVTH-----NOEVSLVKEPIIKOTQASSSISG-ADY 47
Db 717 ETVKLSTIENKKNELLSIIVEMKKHHSBLNELNKIVDFPKSKQLSSNINDYSNY 776

Qy 48 AESGSKSLKINETSGPVDDTVTDLFSKRTTPKIDKNLAKPQELKAVTENTSEK 107
Db 777 KDELNKKYSKISEIKSYND-----QSNIDNIDKEDAK-----QNYEKS 816

Qy 108 QITSGSQLEQSKESLSLNKTPVPSNSWEICFITKGNLTGLSKSGVEKLQTHLVLP 167
Db 817 EYIKTISVKEDEIFKIINE-----MKFMKDILNKNVFNVLNNHKEKIN-----S 863

Qy 168 QAADGTQLIQVAFPAFTPKKTATAEYTSRAGENGESQLQVDGKEIN-----EGRV 221
Db 864 GHSFAELVNKKIENISDQ---LNDYKKNV-----DSKSLINEITKIEEBYQ 910

Qy 222 NSYLLKKVYIPTGYKHGQDAVDNKNIAEVLNPESELETI--SDYAFALHAKQIDLPD 279
Db 911 NINTLKKV---NGYLKICK-----NTTESIEKFRNKQKWLNEILNKNIDVIKN 955

Qy 280 LKATGELAF---FDNQITG-KLSLPQLMLRAEAFKSNHKTIEF-----RGN 324
Db 956 SNLI-EKSYTNQFNDLTLDKXTELEKIFTELSSLYEAKNELIKYFNDLKENLGTGPKGN 1014

Qy 325 SL-KVIGASFDNDLSQ-----LMLPDGLEKI----- 351

```

```

Db 1015 TLVQOPDENEKATNDIEQKSVNANKVSNIEWIHTSIYNIIDIEKLGKIGNIELLKEI 1074
Qy 352 --ESEAFTGNPGD-----DHYN-----NRV-----VLWTKSGKQKPGSLA 383
Db 1075 LKEABISITNFEIKELKHVNPDDFVKEENIKYADBEINKNDIKTLDOQKVDKNKALT 1134
Qy 384 -----TENTYVN-----PKSLWQESP-----EIDYTKWLEEDFTY 414
Db 1135 EIKNKEN-YINEIKTOISDLEVDTKIYNEDPKETIEKIENTIVTKIDKKKIYDNNKK 1193
Qy 415 QKNSVTGF--SNKGLQVKRNKNLEIPKQNGVVTITEIGNAFRNVDFONKTLKY--DL 470
Db 1194 LLNEIAIEKDKTSLEEVK-NINMSYCKSLNKLFLFKI-DEEKKKSENMIKSMKYIKDL 1251
Qy 471 EEVKLPSTIRKIGAFQSNLNKLSFEASDDLEETKEGAFMNNRIETLELKKLVITGDAA 530
Db 1252 DEIKEQSPKAEAMNTFNISNYSKYDHYITS-----QKNDKSIISIDREKSLKTEGN 1301
Qy 531 PHINHIAIVLPESVQEIGRSAPQNGANNLIPMGSKVKTIGEMAFLSNRLHLDLSEQK 590
Db 1302 YKSNINDI--KKTQTYLLDAQKHNSDINLY-----LNEITNLYNLIK----- 1344
Qy 591 QLTEIPVQAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSIAFNALDDNDGD 650
Db 1345 -----NNIKNIIDEVKEYTKIEYNNQNVKSELDKSETL-----IKTIKENSNL 1388
Qy 651 EQFNKVVVTHNSYALADGEHFVDPDKLSSTVDL-EKILKLIBGLDYSTLRQT-TQ 708
Db 1389 ETCKSKI-----ESTVDGK-----DINECICKVTSKKNYILSEESNND 1426
Qy 709 TQPRDM--TTAGKALLSKS-----NLROGEKQKFLQEAQFFLGRV--DLDKATAKA 755
Db 1427 TYFKNAKENNENSVLLFKNIEMANNKTQHIIMONQKNATSDMDPNLNLKENIDKSKKCK 1486
Qy 756 EKALVTTKATKNGOLL-----ERSINKAVLAYNNSAIKKNVKKRLEKELDLLTGLVEGKGP 811
Db 1487 DEADKNAKQTEKNKILFKQYKQKDVTELLNKYSELAIGN-NIAQTKKDSNII----- 1536
Qy 812 LAQATMVQGVVLLKTPLPPEYIIGLVYFDKSGKLIYALDMSDTIGEGQKDAYGNPILN 871
Db 1537 -----INEIKELKKQITL-----QAETSEKINKIKKEK-----T 1567
Qy 872 VDED---NEGVALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAHYHAGIFQAION 928
Db 1568 IEDDAANNKSNKAAIGIOTSLENLE-----NKLKITNIK-----KKIND 1608
Qy 929 AAAAEQQLPKPGTHSEKSSSSSSANSKDRGLQS 962
Db 1609 CLTETESI-----EKQJSSFSINSQDTLSS 1634

RESULT 19
Q7RELO PRELIMINARY; PRT; 2664 AA.
AC Q7RELO;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Rhostry protein (Fragment).
GN Name=PY05054;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=1236865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiolini S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,

```



```
QY 75 -----DKRTPTEKIKDNIAKAPREOEALKAVTENTSEKQITSGSQLEBOSKE----- 120
Db 123 TGRIVYPSSTIPSSIKIKQKGFHSGAKTII-----FDKGSQLEKIEDRAFDFPS 174
QY 121 -----SLSLNK-TVPTSTSNWEICDFITKNTGLVLSKSGVEKLS 158
Db 175 ELEETELPASLEYIGTSAPFSQKLLKLTFFSSSKLEL-----ISHAEAFANLS 222
QY 159 QTDHLVLSQAD-QTQLIOVA-----SFA-----PTPKKTAIABYTGRA 198
Db 223 NLEKLTPLSKVTKLNLNFRLLTSLKHVDVEEGNESFASVDGLVFSKD-KTQLIYPSQ- 280
QY 199 GENGISQLDVGKEIIEGEVFNYSLLKVKVITPTGYKHIGDAPVDNKNIAEVLNPSL 258
Db 281 -KNDESYPKPKTELKSLSYFNKNSY-LKKUELNEGLSKIGTFAPADAKLEELSLPNSL 338
QY 259 ETISDYAF-AHLAKQIDLPNLKAIGE-----LAFDNOQITGKLSPLRQMLRLAE 308
Db 339 ETIERLAFVGNLELAKELILPDNVKFGKVMGNLKPFLTSGNNIN---SLPSPFLSGVL 395
QY 309 RAFKNHJK-----TIEFRGNS--LKVIGEASFO-DNDLSQMLPDLGL 348
Db 396 DSLKEIHIKNKSTEPSVKKDTFAIPETVKFYVTSHEIKDVLKSLNSTSNDI-----IVEKV 451
QY 349 EKIESEATGPGDDHYNRVVLTWK-----SGKNPSGLATENTYVNPDKSLW-- 396
Db 452 DNIKETVAKP-KKNSQGVGVWVKDGLWYLNESGSMATGWV-----KOKGLWY 503
QY 397 -QESPEIDYTKWLEED--FTYQKNS---VTGP-SNKGL 427
Db 504 LNESGSM-ATGWVKDGLWYLNESGSMATGWVKDGL 540
QY Q9V736 PRELIMINARY; PRT: 3257 AA.
AC Q9V736;
DT 01-MAY-2000 (TreeBrel. 13, Created)
DT 01-MAY-2000 (TreeBrel. 13, Last sequence update)
DT 01-MAR-2004 (TreeBrel. 26, Last annotation update)
DE CG12864-PA.
GN ORFNames=CG12864;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

```
RA Merklov G., Milahina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaninker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003814; AAF58230.1; -
DR FlyBase; FBgn0026427; Su(var)2-HP2.
DR GO; GO:0005720; C:nuclear heterochromatin; IDA.
DR GO; GO:0016458; P:gene silencing; IMP.
DR InterPro; IPR000637; A-T hook.
DR Pfam; PF02178; A-T hook; 1.
DR SMART; SM00384; A-T hook; 2.
SQ SEQUENCE 3257 AA; 355970 MW; 7E9212C20A040810C CRC64;
Query Match 4.3%; Score 220.5; DB 2; Length 3257;
Best Local Similarity 18.9%; Pred. No. 0.87; Mismatches 437; Gaps 51;
Matches 221; Conservative 185;
QY 2 KKHLLKTLVTLTTSVTVTH-----NQEVFLSVKEPI--LKQTOAS-----SSISGADYAES 50
Db 1206 KKEAVVGPDLTKTSTSSNTIIDKKSNSFDSMQPSDRLNQKESAFTKLSSISSPKKIMK 1265
```

QY 51 SGKSKLKINETGPDVDTVDLFSKRTTPEKIKONLAKGPRE--OELKAVTENTESEKQ 108
Db 1266 DQDKOLDALSKGDSNPTIRDTGDSRQTDKKHQENDTKHEBEDSSKLKANIDETKSSSE 1325
QY 109 -----ITSGSQLEQSKESISLANKTVSTSNWELCDPITKGNLTIVGLSKSGVEKLSQTDHL 163
Db 1326 KDAEPIKSDOSSQDSAPKPRISKPSKRNKRKNNE-----KKPND--STAESDIEGGFQVN-- 1376
QY 164 VLPSQAADQTLQIVASFAFTP---DKTAIAEYTSRAGENGIEISOLDVDGKEIINEGBV 220
Db 1377 -----TETVQAT--CSTSESNNKDMVK--SDETNEEPLNSETIG--RIRKGOA 1421
QY 221 FNSYLLKKVITPGYKHGQDAFVDNKNIAEVLNPSLETISDYAFAPHLALKQIDLPDL 280
Db 1422 FH-----IENPKDDLHITPON--ENQSIAGVNF-----KQVLPESV 1457
QY 281 KAIGELAFDNIQTKLSLPRQLMLRAERAFKSNHIKTIEFRGNSLKVIGEASFQNDULS 340
Db 1458 ES-----DTPI--MKIPTKTYLMCTK-----NKTSL--SASEDDPDI- 1490
QY 341 QLMPLDGLKIESEAFNGPDHNNRVVLMTKSGKNP----- 379
Db 1491 --VLEPOK--ITTSKGDSPDLNANN-----LETSTQDPKEHEFSQDTTDSNDIPST 1544
QY 380 -----SGLATENTYVNDKSLWQSPEDYTKWLEEDFTYQKNSVTFGSKNG 426
Db 1545 KKSQIVFPTPTTKSSDOTKNSFITPNRSPKSNVSKAKRLDNPFEESQNAASESSASK 1604
QY 427 LQK-----VRNKNLEIPKOHNGVITTEIGDNAPRVDPQNTLRYKLDL 471
Db 1605 VQKELRTPTASCRKLRLVLRKTPTSLSL-----TNSRKSTFKKTPAKSRRLTKILES 1656
QY 472 EVKLPS-----TIRKIGAF-----FOSNLSKFEASDLEIEKEGAF 509
Db 1657 MEKTPRESVSLGEVNDPDPVAESVAVLHSDRDLSENEIPNEEVEFDEE--ASABD 1715
QY 510 MNNRIETLKLKLVGTIGDAAFHINHIYAVLPESVQETGRSAFRONGANNILFMGSKVK 569
Db 1716 TDKLKKKE-----DDEHELVNDICAASNKPTDDSTKQASSNSTDSVLQETKDE 1767
QY 570 TLGEMAFNLRLHLDLSQKQLTEIPVOAFSDNALKEVLLPASLKTIREEAFKNNHLKQ 629
Db 1768 -----LNSLNATQGEDTPIKELTEBEVNN-----KTVEDESKQBEILKQ 1809
QY 630 LEVASALSHIAPNALDDNGD-----EQFDNKVVVKTTHNSVALADGEHFIIDPKLS-S 683
Db 1810 LEPDNA-----ALEEDTASTAKAAEEMDLTYKEKSNVKS-VLAEPETDVTDDDELAQS 1861
QY 684 TIVDLEKILKLEGLDYST---LROTTOTQFRDMTTAGKALLSKNLRO-----GEKQ 733
Db 1862 PIPNSETTSVTDPPSTSSVVKSLRREADSSQPDDEA--AKRKQROVEKSLTGKKE 1919
QY 734 KFLQEAQFGLGRVLDLKAIAKAEKALVTYKATNGQLLERSINKAVLAYNNSAIKKAN-- 791
Db 1920 QVKPARRRQLAEVE--ERPSLKRKSTSEAKSTVGKYISIIQNETIMSTTAPIRETNR 1978
QY 792 -----VRLKELDLDTG-----LVEGKGPLAQATWQGVYLL 824
Db 1979 AASTSPSARKSAVQEAHVETTKTHIILGPPGKLLHSDSPAAEVKPMVQ--TLLSSTLSL 2037
QY 825 KTLPLPEYIIGLVYFDKGLIYALDMSDTIGEGQKDAYGNPILNVDEEDNEGYHALAV 884
Db 2038 QKPSLT-----DGSPKLRKSLKKSIADE-----NIDGQSIFSSSSV 2076
QY 885 ATLADYEGLDIKTILNSKLSQTS--IRQVPTAAVHRAGIFQAIQNAABAEQQLPKPGTH 943
Db 2077 --LNKNTSVVARKVNIIVSLQSKDQVETAA-----SSSTPILTKEKLLK 2122
QY 944 SKGSSSESAN-----SKDRGLQSNP--KTNRGHRSAILPR-----TGSKGSFVYGI 988
Db 2123 TQKSTKPKGNKTESKKSLSLVQGPQMTOKSEAVSGPKILNKYLKSETSSSRKTVSTV 2182

QY 989 LGYTSVALL-----SLITAIIKKK 1007
Db 2183 TGRKQIGQLEVLKPKPSRKSESLVEAISRK 2214
RESULT 22
QYVJQO PRELIMINARY; PRT; 1443 AA.
AC Q9VJQ0; Q9NK96;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE CG4168-PA.
GN ORFNames=CG4168;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang M., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Frannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler P., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=2426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]

Db 1491 SKLDQEIKTKEAEENLNTNAKTELSNLTQAPAEAIKSAKKVLEELINGLNSNDSSIK 1550
 Qy 702 TLRTQTOQPRDMITAGKALLSKSLNRQEQKQFLQBAQFFLGRVLDK-----AIAKAE 756
 Db 1551 SLKATO-KIRDAETQLTAEIKT-----EKLKFEDEARKSLRELKIDDDAKREVETSKAK 1604
 Qy 757 KALVTAKTKNGQLLERSINKAVLAYNNSAIKCANVKLEKELDLTLGLVGGKGPLAQAT 816
 Db 1605 BALKTKKADENSSIEE--IIKATYALEDAKFKELNQEIYVK--DRLMKLNAKKNKLLDA- 1659
 Qy 817 MVQGYLLKT-----PLPLPEYIYGLNVYDFSGKLIYAL-DMSDTIGBQKDAYGNPILN 871
 Db 1660 -----LLKTNLDDLMSIEKLGSL-----AGEIKNKLEDAKLLQEAAGLSKITSK 1708
 Qy 872 VDENEHGHALA-----VATLADYEGLDIYKILNSKLSQLTSTR 910
 Db 1709 IKOKIQIKIDYVROPEPWISNMLKIKG-GLSVINSNIEQLKKPK 1752
 RESULT 25
 PCP1_SCHPO STANDARD; PRT; 1208 AA.
 AC Q92351;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Spindle pole body protein pcpl.
 GN Name=pcpl; ORFNames=SPAC6G9.06c;
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=21852775; PubMed=11864908;
 RA Flory M.R., Morpew M., Joseph J.D., Means A.R., Davis T.N.;
 RT "Pcp1, a Spcl10p-related calmodulin target at the centrosome of the
 RT fission yeast Schizosaccharomycetes pombe.";
 RL Cell Growth Differ. 13:47-58(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne K., Lyne K., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jørgensen K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Weijters I., Vanterreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux C., Lelaie V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomycetes pombe.";
 RL Nature 415:871-880(2002).
 CC -!- FUNCTION: Spindle pole body component that binds calmodulin.
 CC Overexpression of pcpl causes the formation of supernumerary SPB-
 CC like structures and disrupts both mitotic spindle assembly and

CC chromosome segregation.
 CC -!- SUBCELLULAR LOCATION: Spindle pole body.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; Z81317; CAB03608.1; --
 DR EMBL; AF348506; AAK31344.1; --
 DR PIR; T39068; T39068.
 DR HSSP; P08670; IGK4.
 DR GeneDB SPombe; SPAC6G9.06c; --
 KW Calmodulin-binding; Coiled coil.
 FT DOMAIN 151 375 Coiled coil (Potential).
 FT DOMAIN 387 803 Coiled coil (Potential).
 FT DOMAIN 874 1091 Coiled coil (Potential).
 FT DOMAIN 1177 1204 Coiled coil (Potential).
 SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADD42424 CRC64;
 Query Match 4.3%; Score 219; DB 1; Length 1208;
 Best Local Similarity 19.8%; Pred. No. 0.25; Indels 278; Gaps 50;
 Matches 227; Conservative 196; Mismatches 443;
 Qy 32 ILKO-TQASSISGADYAEASGSKLKNITSGPVDVTDVDFSDKRTT---PEKIDNL 87
 Db 104 ISKQATQEALSISQGN--DSVDSKLTDLKSNSEIDHTDGLPANAALTREQKVLK 161
 Qy 88 AK---GPR-----EQLKAVTENTSEKQITSGSQSLESLSINKVTP 129
 Db 162 SRENFGRLRIKIVCLEKLESMAPEQIKAEVKDNVELHAE-RANLQQLKRTESLQKS-- 218
 Qy 130 STSNWEI---CDFITKGNLT-----VGLSGVKEKLSQTDHVLPSQAADGTQ 174
 Db 219 EDKMFLEKVDYLSKNDVEQSONVKVFTIRFLENALKVQREKDSISTEMEED--- 275
 Qy 175 LIQVASFAPTPDKKTAJAEYTSRAGEN--GEIS-QLDVGDKKEIINEGEVFNLSLLKVTI 231
 Db 276 -----KSNKEVDVEYEQRLQNLRLDELSELVDV-AQDLLTEKEDETATLKQIE- 323
 Qy 232 PTGYKHIGQAFVDNKNIAEVLNPLESLETISDYAFALHAKQIDLPNLKAI-CEL-AFF 289
 Db 324 ---EKENSsafeneenssyvhlqe-----DYAI--LQAKCDEFADRIQVLTADLEK 372
 Qy 290 DNQITGK-----LSLPRQLMLRAERAPKSNHKTIEFRGNLSKVIGE--ASFQNDLSQ 341
 Db 373 ENQIMHSEASIGLTDMSQVHTLQEQHLKAN--BEIEFLHDIQSRMNEEGKNFEDIMLQF 430
 Qy 342 LMLPDGLKEIESEAFNGDDHNNRVLVMTKS--GNPGLATENTYVNPDKSLWQESP 400
 Db 431 RSLSEERDVLESKLQTLLE--DD--NNSLRLTWSSLSGNQIESLRTQNREIDEEK----- 479
 Qy 401 EIDYTKMLBEDFTYQKNSVTFGSKGLQKVKKNKLEIPKQNGVTTIEGDNAFRN--- 457
 Db 480 --NHLRLLA-----SKNSKALAEATNIRLOEVTKETLRMKNNDLNEIHLRENEGL 532
 Qy 458 -----VDFQNKTLRYKDYLEVKLPSTIRK-----TGAFQ 488
 Db 533 TLKIDSTIKEDRLINLEBQRIKSYFVNVSELNCTIDBYRNKLKDKETYNVWNAFOYK 592
 Qy 489 SNNLKSFEASDDLEIEKEGAFPMNNRIETLEKOKLVITGDAAFHINHIYVLP--ESVQ 546
 Db 593 DNDLRRPHESINKLQDREKELTSN---LEKNLVIS-----SLRETVAMLEKERSIK 642
 Qy 547 E-IGRSAPRQNGANNLIFMKSQKVTLCGEMAFLSNRHLHDLSEKQKLTPIV-----QAF 600
 Db 643 KYLSGNKADLNTNLMELINDKISVLQKQ--LTDVKDELVDVSEEREEREAIVAGOKLSASF 700
 Qy 601 SDNAKEVLLPASLKTIREAEAFKKNHL--KOLEVASALSALHAFNALDDNQ---DEQFDN 655

```

Db 701 ELMSEKQALELKYSLKNEELINAQNLDRREELSELSEKLLFEERKINSGSNDIEKNK 760
QY 656 KVVVTHHNSYALADGEHIVDPDKLSSTIVDL-----EKILKLEGLDYS 701
Db 761 EIVNLSELADKLAQIRHLESQWELDKLVHHLNRGIEBEANVEANVKRCLGLMGCDYS 820
QY 702 -----TLRQTQTQFRDMWT-AGKA----- 720
Db 821 SVSILQIVSQIEHFVNQQTIRSLKQLRHDFVQPSGKQEQLSRSPEKFLGTETKHD 880
QY 721 LLSKSNLRGEGKQKLOE-AQFELGRVD-----LDKAIKAEKALVTK 762
Db 881 ILAQRNRNVSEKNDLENAQKFFSPDRKNGYLPSEHTSKIEYLETIEDLKAL--Q 938
QY 763 KATKQQLERINKAVLAYNNSAIK-KANVRLEKELDLLGLVEGKGPLAQATWVGQV 821
Db 939 DELKRNLLMDIIS-----SYNKQTYKQEKIKWLERERSILLDELES----- 981
QY 822 YLLKTLPLPEYVYIGLVYFDSKGLIYALDMSDTIGEQQ--DAYGNPILNVDEDEGY 879
Db 982 -----YRSNQFNYQNNLVQDKN-----ELEERLKEIQELEVYNNHFMKQAE----- 1023
QY 880 HALAVATLADYSGLDIKTILNSKLSQTSIROVPTAAHYRAGIFQAIQWAAAEAEQLLPK 939
Db 1024 --LMTSNVTDESQMLKTLREALQSKTNIDHLSLTILERNRKEYKSLLDVYNQLRARYKN 1081
QY 940 POTHSEKSSSSANSKORGLQSNPK-----TWGRHSALTPTGSGKSPVYGILGVTVA 995
Db 1082 LOSNTPQSTQSQYSEIKGLSKLYKLOSKCRREHSLDLDAFKKFKFILMQLTGYTCN 1141
QY 996 LLSL 999
Db 1142 KINL 1145

RESULT 26
Q7RHM8 PRELIMINARY; PRT; 1675 AA.
AC Q7RHM8;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Protein kinase domain, putative.
GN Name=PY03956;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.B.,
RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzbeg S.J., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RL parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519 (2002).
CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABL01001181; EAA15751.1; -.
DR HSSP: O14965; 1MQ4.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.

```

```

DR GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR011591; Botulinum.
DR InterPro: IPR011009; Kinase like.
DR InterPro: IPR000719; Prot.Kinase.
DR InterPro: IPR008271; Ser_thr_pkin_AS.
DR Pfam: PF00069; Pkinase; 1.
DR ProDom: PD001963; Botulinum; 1.
DR ProDom: PD000001; Prot.kinase; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; UNKNOWN_1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1675 AA; 195244 MW; 281BE2B42E5DD60 CRC64;

Query Match 4.3%; Score 218.5; DB 2; Length 1675;
Best Local Similarity 19.2%; Pred. No. 0.42;
Matches 228; Conservative 192; Mismatches 466; Indels 303; Gaps 51;

QY 21 NOEVPSPVKEPIK---QTQASSISG-----ADYAESGSKSLKINETSGPVDVDTVD 71
Db 18 NKNPSNLVNDILKFNFTWI KRNNVVERFVHNKVSDDSNMYNKNKGHHINKSVG-----AEKN 73
QY 72 LFSDKRTTPEKIKONLAKGPREQELKAVTENTSEKQITSGSQLESLSLNTKTVST 131
Db 74 ISGNKTTNLVNIKN-----KNQRLFSKEN---EKVIRINRMREDNYTTNLIK----- 119
QY 132 SNWEICDPFITKGNLTVLGLSKSGVEKLSQTDHLVLPQAADGTOLIQVASEAFTPKTAT 191
Db 120 -----GKNDIINNRKN--TWNNVEYLNLSKSDDKLTNINTSGYAT-----ENI 163
QY 192 AEYTSRAGEGEISQLDVDGKEIINEGEVFNFSYLLKVTIPTGYKHIGQDAF-----VD 245
Db 164 LEDNCNNGNDIRNDIEKDNNGNNGYYSCKNLKKVKEPNN-----SLDSPYKRRKRETE 219
QY 246 NKNTAEVNLPESE-----TISDYAPAHALAKQIDLPNLKALGELAFPDNQITKGL 297
Db 220 EENYDKKRINHSKNDYYNTYTKTDINDSHMNIKDIQIKOISENVSKIKNLENNIMNVK 279
QY 298 SLPRQLMELAEAPKPSNHIKTI--EFGNSLKVIGEASFQDN-----D 338
Db 280 NWEKARSVHNISESEKKFNIIINQSSNNSTKILNSHFSNTNESCNVYKMKDIKNKVD 339
QY 339 LSQMLPDG--LE-----KIESAFTGNPGDDH-----YNNRV---VLWTKSGK 377
Db 340 YKKIKNPENSYLETDENIISRSYKLEEEHI--NNSDDFIRNFWYDNKLETKINETKSK 397
QY 378 NPSGLATE-----NTYVNPDKSLWQESP-----EIDYTKW-----LEEDFTYQK 416
Db 398 LEKNILTDIKSKNSNEYILKHKYNEDEKGNFCDNEFKQNKYDFISCSFNEEEEFMNDY 457
QY 417 NSVTGP---SNKGLOK-VKRNKNLE-----IPKHNGVITIETGDNAPRN----- 457
Db 458 INNTDFVKNRKEGLNKLIDINKRBEQNAAFILKNKNNINNETOBYNNSHENIKULLYNNV 517
QY 458 -----VDFQNKTLRKYLDEEVKLPSTIRKIGAFQAFQSNLKSFEASDLEETKEGAF--- 509
Db 518 KDANEIEFNDISLNDPKMDVSKPIDTKNTKISQYQESNII--YKINNDNEAKDNWEYD 575
QY 510 ---MNNRIETLEBKLVITIGDAAPHINHIYAIVPESVQEI---GRSAFRONGAN----- 559
Db 576 LSSNN---CKNQNDGLKTIIG-ALANDIMNTYKRLVDNTNNVLNKNQKSDSESNLANKK 631
QY 560 -----NLIIFWG-----SKVKTLL--GEMAFSLNLEHL 585
Db 632 RTEKKNALDHKNRVERNRLIISASKNNTNNSNVSVKRVNNSGDISSTSENKKNL 691
QY 586 LSEKQKLTPIVQAF-----SDNALKEVLIPASLKITREAEFAKGNHLKQLEV 632
Db 692 IKKTYNESISKRLFSTHTLSSNNKCLSSGNYLKNKIKP-SIKNSFKTYKKNLDRHKT 750
QY 633 A-----SALSHIAFNALDDNDGDEQFQFNKVVVTKTHNSYALADGEHFVDPDKLSSTI- 685
Db 751 IGFNRNTNLDKINSNISQNNGITKNSIQKPPSSNIHTNLSNIQSKKKTIDNNYKNSTIE 810

```



```

QY 686 -VDLEKILKLEGLDYSLRQTOTQPRDMTTAGKALLSKSNLRQ---GEKQKFLQEAQF 741
DQ 811 NKNLEENNEKIKKQKIBFNKYNNSFLER-----MKKKNLNEHVDGNKNAFTKQTVQ 863
QY 742 FLGRVLDLKAATAKA-----EKALVTCKATKNGQLLE-----RSINKAVL--AYNNSAI 787
DQ 864 NNNINFEKNTKMPKPLSNKKITNNGYKNGKGLFSELAEKKISKSTDKLTIDKKYNHSGH 923
QY 788 KKANVAKRLEKELDLTLGLVEGKGPLAQATQGVVLLKTLPLPEYIYIGLVAYFDKSGKL 847
DQ 924 VEVGAK-LEKGL-----GK-----EIEGLQGEISINSKV 952
QY 848 IYALD-----MSDTTGEQKDAYGNPILNVDEBNEYHALAVATLADYBGLDIKTLNLSKL 903
DQ 953 EEKIEDPLYKQNNVGE-----VFRENMDTNEISNINYSKSNTRKAKANALNRERINNNK- 1007
QY 904 SOLTSIRQVPTAAVHRAGIFQAIQNAAEAEQLLPKPGTHSEKSSSSSESANSKORGLOSN 963
DQ 1008 -----BIRTLASSNS-ITSGKGNLTLVKGQNTNNYMYRYSKKEIEDNKLKMKNLANN 1059
QY 964 PKTNRGRSAIPL-----RTSGKGSFVYGIILGYTSVALLSLITAIKKKK 1007
DQ 1060 ---NKGIDNKYVPKSSNKTSSSSNITNNIVDKKNMLVKKDEINKKKK 1105

RESULT 27
Q6K1F4
ID Q6K1F4 PRELIMINARY; PRT; 662 AA.
AC Q6K1F4
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Putative expressed membrane/lipoprotein.
GN OrderedLocusNames=MMOB1360;
OS Mycoplasma mobile.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2118;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=163K / ATCC 43663;
RX PubMed=15289470; DOI=10.1101/gr.2674004;
RA Jaffe J.D., Stange-Thomann N., Smith C., Decaprio D., Fisher S.,
RA Butler J., Calvo S., Elkins T., Fitzgerald M.G., Hafez N.,
RA Kodira C.D., Major J., Wang S., Wilkinson J., Nicol R., Nusbaum C.,
RA Birren B., Berg H.C., Church G.M.;
RT "The complete genome and proteome of Mycoplasma mobile.";
RL Genome Res. 14:1447-1461(2004).
DR EMBL; AE017308; AAT27622.1; -.
DR InterPro; IPR007326; Lipoprotein_17.
DR Pfam; PF04206; Lipoprotein_17; 2.
KW Complete proteome; Lipoprotein.
SQ SEQUENCE 662 AA; 72407 MW; BB9543519515C79D CRC64;

Query Match 4.3%; Score 218; DB 2; Length 662;
Best Local Similarity 20.6%; Pred. No. 0.12;
Matches 159; Conservative 130; Mismatches 261; Indels 220; Gaps 42;

QY 123 SLNKTVPSTNWEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQAAADGTOLIQVASFA 182
DQ 19 SLTITIVSCSTASISDL-----QLV-ISKNSKITASQNVPASSITSD-EMILQINLKL 71
QY 183 FTPDKKT---AIAEYTSRAGENGETSOLDVQKEITINEGEVFNPSYLLKKVTIP-TGYKHI 238
DQ 72 TIPSVSTNINLIFQLVNGNANDQTSGLD1--KIIATNGTVGNQISIESQTLTVTGFLTQ 129
QY 239 QGDAPVDNKNIAEV-NLPSELETISDYAFAPHALKQIDLDPNLKKAIGELAFPDNQITKGL 297
DQ 130 TQ-----VNNTNLEIKNIFDRFETSLKGVSGIVSARTQITN-----ETNPVIGSN 177
QY 298 SLPQLMRLAERAPKSNHIKIT-EPR-----GNSLKVIGEASFQNDLSQLMLPDGLSKI 351
DQ 178 TLTP-----SKVSQSLNLTILQGLQEPQSLNSASNSVNLTEANYIDGSAN-----DALGTI 229

```

```

QY 352 ESEAFTCNPGDDHYNRRVVLWTKSGKNP-----SGLATENTYVNPDKSLMDESPBIDYTKW 407
DQ 230 QIS-----LQGTIAGNIPTISKSKSVTIGGFLKAEKQILLESNK-----NW 270
QY 408 LBEEDFTYQKNSVTGFSNKG-----LQVKENKNLEIPKQ-HNGVTITIEIGDNA 454
DQ 271 IRD---ISKETIONFANSNKEAFINDQGKULLSTSLKNRNFSLRSSIKNOLNVNYSGPS 327
QY 455 FRNV-----DFQNKTRKRYDLEEVKLPSTI--RKIGAFAPQSN 490
DQ 328 FFNTIINSNODYSIVATISGENISNDFEGSVLDFQI-----TIGNREIGTF-PSSS 379
QY 491 -----NLKSPASDDLEIEIKEGAPMNNRIETLELKD-----LVTTIGDAAFHIN----- 534
DQ 380 LIEKVVNLIGFKKSSSLH-----NTIFVRDKFNVRSLSVLNTGEISAEELNGPDTN 427
QY 535 -----HIYAIVLPESVQIEGRSAFRGANNLIPMGSKVKTLGEMAFSLNRLHDLSE 588
DQ 428 LNPFILEDLTRIVIPSTVEDIGGNIP-----KVKSIKDSAPINSNLRLLDLSN 474
QY 589 QKQUTEIPVOAFSDNALK-EVLLPASLKTIREAFKKNHLKQLEVASA-----LSHIA 640
DQ 475 ASNLESIGNFSPSNKNIENSIQIPASLKNIGENAFSNLNKNINFGNASLSOLEKINKAA 534
QY 641 FNALDDNDGDEQDFDNKVVVKT-----HNSYALADGEHFIVDDPKLSSTIVDLEKILKLE 696
DQ 535 FL-----NNNLESFQVNGVFSNLPFIGNAFANGNPEN-----SRLST--ISLDSIINFLN 583
QY 697 GLDYSLRQTOTQPRDMTTAGKALLSKSNLRQGEKQKFLQE--AQFPLGRVDLDLKAIAK 754
DQ 584 DFTFT-----QDLTI-----EKGAFLRELKSDFWIPQYNLN----- 614
QY 755 AEKALVTCKATKNG-QLLERSINKAV-LAYNNSAIK--KANV-KRLEKEL 799
DQ 615 ---ALNRLHVLKNSIKFLSNQNNNGIDTSPGNVAVSNPKTNIQEKYEKNI 661

RESULT 28
Q7RR45
ID Q7RR45 PRELIMINARY; PRT; 2341 AA.
AC Q7RR45
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY00892;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.B., Riedmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabibi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000239; EAA19159.1; -.
DR InterPro; IPR003900; KID_repeat.
DR Pfam; PF02524; KID; 9.

```


KW Hypothetical protein.
SQ SEQUENCE 2341 AA; 266712 MW; 51D2041A084F4BA5 CRC64;

Query Match 4.3%; Score 218; DB 2; Length 2341;
Best Local Similarity 21.1%; Pred. No. 0.71;
Matches 207; Conservative 147; Mismatches 396; Indels 232; Gaps 46;

QY 53 KSKLKINETSGPVDVTDTLDFSDKRTTPBKIKDNLAAGPREOEL-----KAVTEN----- 102
DB 845 KTELANEHISKKWENTINNEKVSNSISEQSDNATKTENELVLKVSKEISSENQKQN 904
QY 103 -----TESEKQITSGS-----QLEQKESLSLNTKTPVSTNSWEICDFYKGNLTVGLSKS 152
DB 905 PTIITOSKDPNKGSIINIYCDYHENEOKSANSADNSSEKE-----KKTNYLINS-- 957
QY 153 GVEKLSOTDHLVLPQAADGTQLIQVAFPT-----PDKTAIAEYTSRAGENGESQL 207
DB 958 --EKLSENEKLNIDNSKNRANICITNSGDHNNKNSNSTYINGSENS----- 1010
QY 208 DVDGKEIINEGEVF--NSYLLAKKVTIPTGYKHIGQDAFVDNKNIAE-----VNLPSLETI 261
DB 1011 --SSKEASKLGSYFKDSEYSMREESEKTNL-HVNDSEYTSIKMSRMASYINGSEK---- 1063
QY 262 SDYAFALHAKOIDLPDNLKALGELAFQNTGKLSLPRQLMLAERAFKSNHKTIFP 321
DB 1064 -----TLSKKDLKINSR-----IDKSEKALNKKDIEIDSH 1093
QY 322 RGNLSLVIGEASFQDNDLSQMLPDGLEKIESEAFPTGNPDHNNRVVLTGSKG--NP 379
DB 1094 INKSEKILSEGGFEIN--SHINNP-----EKWANE--KGLSIDSHLN-----KSGRMVNE 1139
QY 380 SGLATENTYVNPDKSLWQSPSEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIP 439
DB 1140 KDLSDSHLNSKGMVNEKDELD-----SHLNSK--CKWNEKDELD 1181
QY 440 KOHN--GVTTIEGDNAPRVDPQNTLRKYDLEEVKLPSTIRKIGAFQSNLNKSPFA 497
DB 1182 SHLNSKGMLEKGLSDNSKSEKTLNKKGLE--INSHISKSG--KTVNEKDLBS 1234
QY 498 SDDLBEIKEGAPNNRIETLELKDLVTIGDPAFHINHIYAIVLPESVQEIERSAPRQNG 557
DB 1235 DSHLN--KSGKIVNEK--GLEIDSHLNSKSGK-----MLNEKGLLEI--DSHLNNS 1277
QY 558 ANNLFMGSKVKTGEMAFSLNRLEHLDLSEOKLTETPVOAFPSDNALKEVLLPASLKI 617
DB 1278 ENTLSKKGLKI-----NSRINNSEKALNKKOLEI-----DSHINK-----SEKIL 1317
QY 618 REEAFKCN-----HLKQLEVASALSHIAFNALDNDGDGEQPDNKKVVKTHNSYA 667
DB 1318 NEEGLKINSRINNSEKTLNKKDPEIDSHL-----NKGSKMVNEKGLESDSH--- 1363
QY 668 LADGEHPIVDPD-KLSSTIVDLEKILKLEGLDYSTLRTQTOTQFDMTTAGALLSKSN 726
DB 1364 LNSEKMLSEKGLKIDNSISKEKTLNK-KGLEIDSNINSEN-----TLSSKGL--KIN 1415
QY 727 LRQGEKQKFLQBAQFPLGRVDLDKATAKAEKALVTKATK-----NGOLLERSINKAVLAY 782
DB 1416 SRINNSEKALNKD-----LEIDSHINKSEKTLNQKLEMAHINGS--EKILSEKGLKI 1468
QY 783 NNSAIKANVXBLEKELDLTLGLVEGKGPLAQATMVQGVYLLKTLPLPLPEYIIGLVNVPD 842
DB 1469 -NSRINNSEKMLSEKGLKIDNSISKEKTLNKKGLEIDSHLNSKGMVNEKGLEIDSHLN 1527
QY 843 KSGKLIYALDMSDTTIGEGQKDAY-GNP--ILNVEDNEGYHALAVATLADYGLDKITIL 899
DB 1528 KSGKMVNEKDL-----EMDHSINPEKILKEGGIEINSHINNSEKMLSEKGLEIDSHL 1580
QY 900 NKKLSQLTIRQVPTAAYHRAGIFQAIQNAAEAEOLLPKPQTHSEKSSSSSEANSKORG 959
DB 1581 N-KSGKILSEKGL-----AIDSHLNSKGMKILSEKGLAIDSHLNSKSEKTPNERG 1627
QY 960 LQSNPKNTNRGRHSAILPRTGSK 981

DB 1628 IEINSHINKS--EKILSEKGLK 1647

RESULT 29
Q9CF64
ID Q9CF64 PRELIMINARY; PRT; 1072 AA.
AC Q9CF64;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein YqG.
GN Name=yqG; OrderedLocusNames=LL1617;
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
NCBI TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471; DOI=10.1101/gr-1697R;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarre K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis il1403";
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006392; AAK05715.1; -.
DR PIR; A86827; A86827.
KW Complete proteome: Hypothetical protein.
SQ SEQUENCE 1072 AA; 113056 MW; 464446E2656CAA08 CRC64;

Query Match 4.3%; Score 217.5; DB 2; Length 1072;
Best Local Similarity 16.9%; Pred. No. 0.25;
Matches 151; Conservative 159; Mismatches 365; Indels 219; Gaps 29;

QY 149 LSKSVEKLSQT-----DHLVLPQAADGT-----OLIQVASFAFTPDKKT 189
DB 1 MSQSKIKITMTVGGLLTATLNPFTLNHQLTELPEATTSKITTYTDEQKITNLDKVS 60
QY 190 AIAEYTSRAGENG-EISQLDYDVGKEIINEG-----VFNSVLLKKVTIPTGYKHI 238
DB 61 LTAEVASVAGLAGADIVIPDI-----IVNNGQTYAITSIGTYAFSGSIRSVIIGNVVDI 116
QY 239 QGDADFV-----NKNIAEYVNPESLETISDYAFALHAKQIDLPNLKAI GELAPFD 290
DB 117 NTSAFQTTTAPPDYVKSLTKVLGAKVQNIKTDAGNAISSIEFPNSVLKIATRAPAN 176
QY 291 NOITGKLSLPQLMLAERAFKSNHKTIEPRGNSLKVIGESFQDNDLSQMLPDGLEK 350
DB 177 NNLT-ELSLGSGNITEIMAKAFQSNQITTIETFADESLLTIVDSAAFSGSVQSLTLGIV-- 233
QY 351 IESRAFTGNPGDDHNNRVVLTGSKG-----NPSGLATENTYVNPDKSLWQES 399
DB 234 -----TLADDVFNKTSPLFGQJSDLPTEIRTIYSNSSL-PDKSWINSDSS---QS 281
QY 400 PEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPKQHGNTI-----TEIGDN 453
DB 282 TSVS-----TENADTTPVSSNSVEQVDVASS-ESTQDANSASLYPISEASSVTDN 331
QY 454 AFRVNDPONTLRKYDLEEVKLPSTIRKIGAFQSNLNKKSFEAD-----DLEE 503
DB 332 TLNLSISLSDSSISSQSTENSO--SGASSTAEISYDSSENSLSLSSNQINSNSSEKDSNQ 389
QY 504 IKGAFMNNRLETTLEKDLKLVIGDPAFHINHIYAIVLP--ESVQEIERSAFRONGANN 560
DB 390 SLLGSSMSSESEHSNSNINETNNSSEITNLPSPNPTESNSVSDQTSSEASTNSNS 449
QY 561 LIFMGSKVKTGEMAFSLNRLEHLDLSEOKLTETPVOAFSDNALKEVLLPASLKTIREE 620
DB 450 ISLSPSNISSTSDSSATNSDPSNVAEVANNLSASVNSSSVLSSTSTADNL-GINQS 508
QY 621 AFKNHKLQLEVASALSHIAFNALDNDGDGEQPDNKKVVKTHNSYALADGEHFTVDPDK 680
DB 509 GSDNLTQDSSEISTSGAFLLSNQTSSE-----ASTNSNS-----SISLSPSN 550

Db	821	-YYINSISIEELDNLDW----	EFYQEDIEDDLEYNFVQESNDWYSDQ-----	E 867
Qy	435	NLEIPRQHNGVTTITEIGD-NAFNRVDFQNKLRKYDLEEVKLPSTIRKIKAGAFQSGNNLK	493	
Db	868	NWYSDQYLGIYMF-IGETSGENNQDNMGEIKEYNKT-----PEKINKI----	IFTSEKFK 920	
Qy	494	SFEASDDLEIIEKGAFAWNRRIETL-----ELDKLVITIGDAAAFHNHYY-----	AIPLPESV 545	
Db	921	QI-MENDLANMTWKDI FHNNKLLFKGKIEIYVITDIEHPIDVQVQNDKIINPLFLYN	979	
Qy	546	QEI-----	GRSAFRONGA- 558	
Db	980	QELFENEIOKIPDKQONKI RNHIAVEIVVKAYFREGIDTPPEIILCDDRTYPOEGSL	1039	
Qy	559	-----NNLIFMGSK-----	VKTLGEMAFLSNRLEHLD 585	
Db	1040	VEVLIGNLVIQVKYFKTIINYSISIEDKNLDSLVMYMNLEGIKMIKDSKIPSIRLRNLY	1099	
Qy	586	LSEQKQLTEIPQAFSDNALKEVLLPASLKTREBAFKQHLKQLEVA-----	SALSH 638	
Db	1100	VLSNKHIVKXKQ-QYNGNI-----IEPIFDQVIQNNRNRYIEYKPGKFDRTKLKS	1150	
Qy	639	IA--FN--ALDNDGDEQFNKVVVKTHN-----	SYALADGHEFIVDPDKLSS 683	
Db	1151	YGRFNEPELRDDRT-NIQREKQDIEKADHNLELQKELNNLNIYSOQG-----	SS 1200	
Qy	684	TIVDLSEKILKIEGLDYSLRTQTOTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFL	743	
Db	1201	NVLIDIPKILKI-----ENTKNYKQFHIIGK-----	ITEGRNLKFPY----- 1237	
Qy	744	GRVLDLKAIAKAEKALVTYKATKNGOLLERSINKAVLAYN-----	783	
Db	1238	--ILIDTG--AADSVISSKILEDEKLVSNKLSKVVTSYNADNKEKHIYDRNTEVIEILID	1292	
Qy	784	NSAIKKNVRLKLEKEDLLTLGLVEGKGLPQAQWTVQGVYLLKT--PLPLPEYVIGLVNY	840	
Db	1293	KNEKYKINPIGLVDQRLLEG-----	GRABILLGMNLIQNLKPYCIITDDYLEINUG 1344	
Qy	841	F 841		
Db	1345	F 1345		
RESULT 32				
ID	005166	PRELIMINARY;	PRT;	708 AA.
AC	005166;			
DT	01-JUL-1997	(TEMBLrel. 04, Created)		
DT	01-JUL-1997	(TEMBLrel. 04, Last sequence update)		
DT	01-MAR-2004	(TEMBLrel. 26, Last annotation update)		
PCPA				
GN	Name=pcpa;			
OS	Streptococcus pneumoniae.			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OC	NCBI_TaxID=1313;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=B6;			
RC	MEDLINE=98340517; PubMed=9675886;			
RA	Sanchez-Beato A., Lopez R., Garcia J.;			
RT	"Molecular characterization of Pcpa: a novel choline-binding protein			
RT	of Streptococcus pneumoniae.";			
RL	FEMS Microbiol. Lett. 164:207-214 (1998).			
DR	EMBL; Z82001; CAB04750.1; -.			
DR	HSSP; P06653; 1H8G.			
DR	InterPro; IPR002479; CW binding.			
DR	InterPro; IPR007093; LRR Tp.			
DR	Pfam; PF01473; CW binding 1; 12.			
DR	SEQUENCE 708 AA; 79360 MW; B7754B63292ECD87 CRC64;			
Query Match 4.3%; Score 217; DB 2; Length 708;				
Best Local Similarity 21.4%; Pred. No. 0.15;				

RESULT 34

GOB1 HUMAN STANDARD; PRT; 3259 AA.
 AC Q14789; Q14398;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgin)
 DE (Golgi complex-associated protein, 372-kDa) (GCP372).
 GN Name:GOLGB1;
 OS Homo sapiens (Human).
 OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN [2]
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94187728; PubMed=7511208;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
 RT protein (giantin).";
 RL Mol. Cell. Biol. 14:2564-2576(1994).
 RN [2]
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94257116; PubMed=8198703; DOI=10.1006/jaut.1994.1006;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as
 RT target of antibodies in patients with rheumatic diseases and HIV
 RT infections.";
 RL J. Autoimmun. 7:67-91(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95100974; PubMed=7802676;
 RA Schda M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of a human 372-kDa protein
 RT localized in the Golgi complex.";
 RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
 CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
 CC of the Golgi complex.
 CC -!- SUBUNIT: Homodimer; disulfide-linked.
 CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.
 CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
 CC autoimmune disease Sjogren's syndrome.
 CC -!- SIMILARITY: Belongs to the golgin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X75304; CAAS3052.1; -;
 DR EMBL; D25542; BAA05025.1; -;
 DR FIR; A56539; A56539.
 DR PIR; I52300; I52300.
 DR Genew; HGNC:4429; GOLGB1.
 DR MIM; 602500; -;
 DR GO; GO:000139; C:Golgi membrane; TAS.
 DR GO; GO:0005795; C:Golgi stack; TAS.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
 KW Antigen; Coiled coil; Golgi stack; Transmembrane.
 FT DOMAIN 1 3235 Cytoplasmic (Potential).
 FT TRANSMEM 3236 3256 Potential.
 FT DOMAIN 3257 3259 Luminal (Potential).
 FT DOMAIN 48 593 Coiled coil (Potential).
 FT DOMAIN 677 1028 Coiled coil (Potential).
 FT DOMAIN 1062 1245 Coiled coil (Potential).

FT DOMAIN 1301 1779 Coiled coil (Potential).
 FT DOMAIN 1828 3185 Coiled coil (Potential).
 FT DOMAIN 2420 2423 Poly-Glu.
 FT DOMAIN 2993 2996 Poly-Ser.
 FT CONFLICT 1 39 Missing (in Ref. 3).
 FT CONFLICT 215 215 A -> AQLSSM (in Ref. 3).
 FT CONFLICT 1765 1765 D -> G (in Ref. 3).
 FT CONFLICT 2950 2950 H -> D (in Ref. 3).
 SQ SEQUENCE 3259 AA; 376075 MM; 60376A20D8A178DD CRC64;
 Query Match 4.2%; Score 214.5; DB 1; Length 3259;
 Best Local Similarity 19.8%; Pred. No. 1.6;
 Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;
 QY 20 HMQEYFSLVKEPIKQTOASSISGADYAESSGSKLKI-----NETSGPVDVTVTL 72
 DB 233 HEDELQLV-----TQAD-----VETEMQKLRVLQRLKEHEESLVGRAQVVDL 277
 QY 73 FSDKRTTPEKIKDNLAKGPREQELK-----AVTENTESEKQITSGSQLEQSKESLSLNK 126
 DB 278 LQOELTAARQNRQILSQQLQQMEAEHNTLRNVTETERESKILLEKMELEVAERKLSPH- 336
 QY 127 TVPSTSNWEICDFITKGNLVLGL-----SKSGVEKLSQTDHLVLPQQAADGTQLIQV 178
 DB 337 NLQEEHMLLEQFEQAGQAQAELESRYSALEOKHAEEMEETSHLSLQKTG---QELQS 393
 QY 179 ASFAFTPPDKKTAIAEVTSRAGNGEI-----SOLDVDGKEIINEGEVFNFSYLLKKVTIPT 233
 DB 394 ACDALKQNSKLLQDKNEQAVOSQTIQOLEDQLOQSKKEI-----SQFLNRLLPQQ 445
 QY 234 GYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFHALLKQIDLPNLKAIGELAPFDNOI 293
 DB 446 -HETASQTSFPDVYN-----EGTQAVTEENIASLQKRVVEL-----ENEK 484
 QY 294 TGKLSLPQLMRL-AERAPKSNHIIKTIIEFRGNSLVKIGEASQDNDLSQMLPDGLEKTE 352
 DB 485 GALLLSISIEBELKAEENEKLSQITLLEAQNR-----GEA-----DREVSIISVDIANKRS 537
 QY 353 SEAFNGNFCDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLMQESPIDYTKMLEEDF 412
 DB 538 SGA-----BESGD-----VLETFPSQKHKL-----SVLLEMKAEQEIEI 573
 QY 413 TYQKNSTVTG-----FSNKLQKQVKNKNLE-----IPKQNGVT 446
 DB 574 AFLKLQLOQKRAEADHEVLQKEMKQMEGEGIAPIKMKVLFEDTGQDFPLMPNESSLP 633
 QY 447 ITEIGNAFNRVDFQNTLRKYDLLE--VKLPSTTRKIGAPAFQNNLKSFRASDDLEI 504
 DB 634 AVE---KEQASTEHQSRTSEELNDAGVELKST-----KQDGKSLSAVDPDQGC 681
 QY 505 KEGAPMNNRIETLELKDKLVTIGDAAFH-INHIYAIVLPESVQIEGR-----SAFRQNGA 558
 DB 682 HQDELERLKSQLEL-----ELNFHKAQEIYEKNLDEKAKEISNLNLTIEEPKKNAD 733
 QY 559 NN-----LIFMGSKVKTIGEMAFISNLEHLDL-----SEQQLTEIPVQAFS 601
 DB 734 NNSAFTALSEERDQLL---SQVKELSMVTELRAQVKLEMLNLAERQRLDYESTAH 790
 QY 602 DNALKEVL---LPASLKTIREAFKQNLKOLEVASALSHIAFNALDDNDGDEQDNK- 656
 DB 791 DNLLTEQIHSLSIIEAKSKDVKIEVL-QNELDDVQLQFSEQSTLIRSL-----QSOLQNK 844
 QY 657 -----VVVKTTHNSVALADGEHFI VDPKLSLSTIVDLKILKLEGLDYSTLR 704
 DB 845 SEVLEGAEVRVRISSKVEELSQALESLEITKMDQLL-----LEK-----KRDVETLQ 893
 QY 705 QTTOFQFDMTWTAGKALLSKNLKQCKQKQFQEAQFGLGRVLDLQAKIAEAKALVTKKA 764
 DB 894 QTIEEKDQVVT-----EISFSMTKXVQLNEEKFSLG-VET-KTLEQNLNLSRAEE 943
 QY 765 TKNGQLLERSINKAVLAVNNSAIKKA---NVKRLKEKELDLLFGLVEGKGPLAQATWQGV 821
 DB 944 AKKEQVEEDNEVSSGLKQNYDEMSPAGQISKEELQHEFDLLKKENEQRKRKLQAALINRK 1003

```

QY 822 YLLKTPPLPEYVIGLVYFDKSGKLIYALDMSDT-IGEGQDAYGNFILNVDBNEGYH 880
Db 1004 ELLQVRSLBELANLK---DESKK---EIPLSETERGEVEE-----DKENKEYS 1047
QY 881 ALAVATLADYEGLDIKTILNSKLSQLTGIR-----QVPTAAVHRAGI 922
Db 1048 EKVTSKQCEIYILKQIYSEKEVLEQIRKDLKLAABEQFQALVKQMNTLQDKTQ 1107
QY 923 FOAIQNAAAEAQLPKPGTHSEKSSSES-ANSKDRGLQGNPKTNRGH 971
Db 1108 IDLLQAEISENQAIQKLTNTDASDGSVALVKETVVISPPCTGSSEH 1157

RESULT 35
Q97NB5 PRELIMINARY; PRT; 621 AA.
AC Q97NB5;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Choline binding protein PcpA.
GN OrderedLocusNames=SP2136;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.B., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007502; AAK76194.1; -.
DR PIR; A95250; A95250.
DR HSP; F06653; IH86.
DR TIGR; SP2136; -.
DR InterPro; IPR002479; CW binding.
DR Pfam; PF01473; CW_binding_1; 7.
KW Complete proteome.
SQ SEQUENCE 621 AA; 69237 MW; 69808PDP43895EB6 CRC64;

Query Match 4.2%; Score 213.5; DB 2; Length 621;
Best Local Similarity 23.2%; Pred. No. 0.18;
Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31;

QY 6 KTVALTTLTVSVV-----THNQEVF-----SLVKEPILK-----Q 35
Db 3 KTTILSLTAAVILAAYVNEPILADTPSEVIKTKVIGSIQQNNIKYKVLTVEGNIGT 62
QY 36 TQASSISGADY-ARESSGK-----SKLINETSGVDVTDVTLFS-----DK 76
Db 63 VOVGNGVTPVEAGQDQKPFTIPKTVGDKVFTVTEVAGQFSYYPDETGRIVYYPSS 122
QY 77 RTTPKIKDNLAKGPREGELKAVTENTESEKQITSGSQLEQSK-----120
Db 123 ITIPSSIKKIQKGFHSGKAKTII-----FDKGSQLEKIEDRAPFSELEIELPAS 174
QY 121 -----SLSLANK-TVPESTSNWICDFITKNTLVGLSKSGVEKLSQTDHLVLP 168
Db 175 LEYIGTSAFSFSQKLKLTFFSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222
QY 169 AAD-OTQLIQVA-----SFA-----PTPKKTAIAEYTSRAGENGISQLD 208

```

```

Db 223 VKTIGSNLPHITTSKLVNDBEGNESFASVDGVLFSD-KTQLIYYSQ--KNDESXTP 279
QY 209 VDGKEIINBEGEVNFSYLLKKVVTPTGYKHHIGQDAFVDNKNIAEVLNLPESLETISDYAF-A 267
Db 280 KETELASYSFNKNSY-LKKLELNEGLEKIGTFAFAIDAILEEISLNSLETIERLAFYG 338
QY 268 HIALKQIDLDPNLKAIGELAP-----FDNQITGKLSLPRQLMLRAEPKFSNHIK 317
Db 339 NLELKEILFDNVKNFGKVMNGLPKLSLTIGNNIN---SLPSFFLSGLVDSLKEIHK 395
QY 318 -----TISFRGNS--LKVTIGEASFQ-DNDLSQLMLPDGLEKTESBAFT 357
Db 396 NKSTEPSVKKDTFAIPETVAFYVTSSEHKVLSNLSSTNDI-----IVEKVDNLIKQETDV 451
QY 358 GNPDDHYNRNVVLWTK-----SGKNPSGLATENTYVNPDKSLW---QSSPRIDY 404
Db 452 AKP-KKNSQGVGVGVKDKGLWYLYNLSGSGMATGMV-----KDKGLWYLYNLSGSM-A 502
QY 405 TKWLEED--FTYQKNS---VTGP-SNKGL 427
Db 503 TGMVKDKGLWYLYNLSGSGMATGMVKDKGL 531

RESULT 36
Q71WB3 PRELIMINARY; PRT; 489 AA.
AC Q71WB3;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Cell wall surface anchor family protein.
GN OrderedLocusNames=LMOF2365_2638;
OS Bacteria monocytes (serotype 4b / strain F2365).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=265669;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15115801; DOI=10.1093/nar/gkh562;
RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,
RA Kolonay J.F., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,
RA Peterson J.D., White O., Nelson W.C., Nierman W.C., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,
RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,
RA Forberger H.A., Tran B., Katharopoulos S., Wonderling L.D., Uhlrich G.A.,
RA Bayles D.O., Luchansky J.B., Fraser C.M.;
RT "Whole genome comparisons of serotype 4b and 1/2a strains of the food-
RT borne pathogen Listeria monocytes reveal new insights into the
RT core genome components of this species."
RL Nucleic Acids Res. 32:2386-2395(2004).
CC -|- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (by similarity).
DR EMBL; AE017331; AAT05403.1; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR InterPro; IPR009459; DUF1085.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR007093; LRR_Tp.
DR Pfam; PF06458; DUF1085; 2.
DR TIGR; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Complete proteome; Peptidoglycan-anchor.
SQ SEQUENCE 489 AA; 53529 MW; 07FA798B2A3B86 CRC64;

Query Match 4.2%; Score 213; DB 2; Length 489;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 118; Conservative 87; Mismatches 193; Indels 164; Gaps 21;

QY 472 EVKLPSIRKIGAPAFQSNLKSFEASDDLEIEIKGAPMNNRIETLKLKLVITIGDAAF 531
Db 51 DITIPITL-----GPNNEYTVTA-----IGNGAFKSKRLTNVITPNTVITIGDAF 96

```


TIGR; EF1824; .
 DR GO:0009986; C:cell surface; IEA.
 DR GO:0016787; F:hydrolase activity; IEA.
 DR GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO:0000272; P:cell adhesion; IEA.
 DR GO:0000272; P:polysaccharide catabolism; IEA.
 DR InterPro: IPR008965; Cellul bind.
 DR InterPro: IPR002105; Dockerin_1.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000421; FAS5 C.
 DR InterPro: IPR011490; FIVAR.
 DR InterPro: IPR003961; FN III.
 DR InterPro: IPR000322; Glyco hydro 31.
 DR InterPro: IPR001899; Gram pos anchor.
 DR Pfam: PF00754; F5_P8_type_C; 1.
 DR Pfam: PF07554; FIVAR; 6.
 DR Pfam: PF00041; fn3; 1.
 DR Pfam: PF01055; Glyco hydro 31; 1.
 DR SMART; SM00060; FN3; 1.
 DR PROSITE; PS00448; CLOS CELLULOSOME RPT; 1.
 DR PROSITE; PS00018; EF HAND; UNKNOWN_1.
 DR PROSITE; PS00022; FAS6C_3; 1.
 DR PROSITE; PS00853; FN3; 1.
 DR PROSITE; PS00847; GRAM POS ANCHORING; 1.
 KW Complete proteome; Hydrolase.
 SQ SEQUENCE 1866 AA; 205565 MW; 4682D5E8B77A96D5 CRC64;

Query Match 4.2%; Score 213; DB 2; Length 1866;
 Best Local Similarity 19.3%; Pred. No. 0.85;
 Matches 239; Conservative 176; Mismatches 423; Indels 398; Gaps 56;

QY 49 ESSGSKLAKINETSQGVDTVTDL-----FSDKRT-----PEKI--- 83
 DB 751 EAQQSATTIN-SQGPKSEKGLDVTIETPKSGYKDFVDERSTTDLALASAEVSTAM 809
 QY 84 ----KDLAKPREQELKAVTENTESEKQI-----TSGSOLESKESLSINKTVPST 131
 DB 810 VGGTEVTLQAANKKEFLAGTNLYYDFKEFQVNOYLSEASGEKLNQSALSVKLAKQSVTA 869
 QY 132 SNWEIC--DFITK-----GNTLVGLSKSGVEKLSQTDHLVPSQAA----- 170
 DB 870 KDQVITVKGFINKGTVDGNGTIV-----DDQLTIPANVAINEEKTTPPSLTILQ 917
 QY 171 -----DGTQLIQVASFAPTPDKKTAIABYTSR---AGENG--EISQ----- 206
 DB 918 WDQVTEATSEYVERDGTFGNIQTNTATDFGFSFLSEHTFRVRANGKGVSEWSEPIKKG 977
 QY 207 -LDVDGKEIINEGEVFNVS-----LLKKVT---IPTGVKHIGQDAFVNKNIAEVLNPE 256
 DB 978 TODDPYKETINQVATSNLPEQGAELKLTDKDLSTGW-----HTNWT 1022
 QY 257 SLETSIDYAFHALK-----QID-----LPDNLKAIGEL-----AFFDN 291
 DB 1023 GIANPSDGNF--LSLKFDLGAEVQMDKIEVLPEDNAGNLIQLQYRTSKDGNWTFESE 1080
 QY 292 QITGKLSLPQMLRLAERAFKSNHIKTIIEFRGN-----SLKVIGEA 332
 DB 1081 PINWKQDALTKTIETKDQAYRFVEMKVLKSVGNFGSGREMLFYKQPGTEGILHGDITNDG 1140
 QY 333 SFQDNLSQLMPLDGLKTESAFTG--NPGD-----DHYNRVVLWTKSG----- 376
 DB 1141 TIDENDAMRYNTGLSEVDSD--FNGYVEKGDUNKNGVDAVDISVLRQLDGGIETIPDV 1199
 QY 377 -----KNPSGLATENTYVNPDKSLMWESPE 401
 DB 1200 EETAGLSLAVNENGKDTYLPDGTITFLKQDLKLNALSTKMSFDSKSELVQGPAT 1259
 QY 402 IDYTKWLEDDFYQKNS-----VTGFSNKGQKVRKNRLKLEIPKHNGVTTIETGDNAP 455
 DB 1260 TNNQTQOMNYSKYRKHSDVENLYLVLSNQG-----NQQL-----LNG----- 1297
 QY 456 RNVDQFNQTLRKVDLEVKLPSTIRKIGAP-----AFQSNNLKSFASDDLEIK 505

DB 1298 -SMDLVTFKVKVKTETTRVKRAATTVPQPLQFDMSQGLLVGQGFQQAATLSDFSVT----- 1349
 QY 506 EGAFNNRIETLELKKK-----LVTIGDAAFHINH-----IYAIVLPVSQVEIGRSAPR 554
 DB 1350 -----VKPTELVDKELLQALITINQARVEKEYTPTWAIKFPILDEAVLANEQAT 1401
 QY 555 QN-----GANNLIFMGSKVKTLGCEMAFLSNRLSHLDLSEQKOLTEIPV--QAFSDNALK-- 606
 DB 1402 QTDVSAAAEENLEKAASOLEKMPDVA--NKAD--LEKATQEGLAQKPSQGEETEETKKVL 1457
 QY 607 EVLLPASLKTITREEAFKKNHLKQLEVASALSALSHIAFNALDDNDGDSQFQDNKVVVKTTHNSY 666
 DB 1458 EESLAAAKVPAQAEKVQTEIDQ-----ATKTLREAL 1489
 QY 667 ALADGEHFIVDPDKLSSTIIVLEKILKLEGLDYSITLQTTQTFQFRDMTTAGKALLSKSN 726
 DB 1490 AOLKEQPVAVDKETLKEQIAQ--ARGKPEEGYQFT---KETEKQLQEAIAQAAEIVAKET 1545
 QY 727 LRQGEKQKFLQEAQOFFLGRVDLDKAIKAEKALVTKKA-----TKNQOLLERS---INKA 778
 DB 1546 ATKEEVSALNALETAMAOQLKEVPLVVKQDQEVVVKRAQQVTPSEGHQTASSLOELQA 1605
 QY 779 VLAYNNSAIKK--ANVXLEKEKELDLTLGLVEGKGLAQATMWQVYLLKTPPLPEY-- 834
 DB 1606 LLAAKNT-LKNPAANQKIMDEAVELTSAIDG---LQEBVLVTDKKALEAMIAKAKIKP 1661
 QY 835 -IGLVNVPDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGHEHALAVATLADYEG 893
 DB 1662 SAGKEFTSESKARLTAIDQABGI--LADKNARQEQI--DIAEKN-----VKTALD--SL 1710
 QY 894 DIKTILNSKLSLTSTSIROVPTAAVHRAG--IPQAIQNAAE-----AEQLLPKPGTHSEKS 947
 DB 1711 E-EQVLQDQTKLKELOKAEETLKPAGKQFTKASQEALEAIKQAKALVEDPNATQEA 1769
 QY 948 SSSESANSK-----DRGLQSNPKTNRGRHSAL----- 974
 DB 1770 DKCLSILQAEAMAEPEISSNSTGNNGHSTVSGTSGQKGTATGTTTKTTTSGT 1829
 QY 975 LPRTSGKGSFVVGILGY---TSVALLSLITAIKXK 1007
 DB 1830 LPKANEVVSPIWSISGFLLIIVSIGLGLFFKNKKEQ 1865

RESULT 40
 Q9AOK5 PRELIMINARY; PRT; 2045 AA.
 AC Q9AOK5;
 DT 01-JUN-2001 (TREMREL. 17, Created)
 DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
 DT 01-JUN-2003 (TREMREL. 24, Last annotation update)
 DE Putative extracellular matrix binding protein.
 GN Name=epf; OrderedLocNames=SPY0737;
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
 RA Perretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon X.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (By similarity).
 DR EMBL; AE006525; AAK33683.1; -.
 DR GO; GO:0009986; C:cell surface; IEA.
 DR GO; GO:0005618; C:cell wall; IEA.
 DR InterPro; IPR011439; DUF1542.


```

Db 1823 -----EGLLKLEDTLRN-----IKLTGTGAPFASL-----SELEKDLIS 1858
QY 472 EVKL-----PSTIRKIGAFAGSNNLKSPFASDLLEI-----KEGAFMNNRIETLELKDKL 523
Db 1859 ENRLAAVDEASVRGIG-----KLREFYINDN-----EIGIDGLSGLDILNRVQTVLSLQ- 1908
QY 524 VTIGDAAAPHINHIYAIVLPESVOEIGRSAPFRONGANNLIFMGSKVKTIGEMAFPLN-PLR 582
Db 1909 -----SIVPDTL-----LANKSMVFYVQGTTPASLPAOFFRSTIALE 1948
QY 583 HLDLSEKQLTEIPVOAF-----LPA-----SLKTIREAFKKNHLK 607
Db 1949 KLSLSDNQLVKLPDFQYVYPLBELVLLNNSLSVLPFGVDPGLIRELDISENPLKS 2008
QY 608 V-----LIPA-----SLKTIREAFKKNHLK 628
Db 2009 LPGLGFAQTFLLESRLMADANLTLTLPAGIFDKLYVLAKLDLANNQLRTLREGVF--NRLY 2066
QY 629 QLEVAS-----ALSHIAPNALDNDGDGEQFDKNVVKVTHNSVALADGEHFIVDPDKL 681
Db 2067 SLETLSENNOLEALQALFKSLEKLN-----IVLSHKNLAAIDPQLFAGLP--- 2114
QY 682 SSTIVDLEKILKLEGLDYSLRTQTQTFQFRDMTTAGKAL-----LSKSNLRQGEKQ 733
Db 2115 -----VTAIDLSNYRLTTFDE--SAMAFAGRLMLMSNESKILTSKIHSTLEK 2160
QY 734 KFLQBAQFPLGRVLDLKAIAKAE-----KALVTKKATKNG--QL-LE 772
Db 2161 LPVDDNKLSSIEVKLEDSYSHITSLSVVRNNFTALDAFQFQMLTELDASYNAPTELNL 2220
QY 773 RSINKAVLAY-----NNSAIKKNVRLK-----ELDLTLGLVEGKGLAQHTWQGYVL 823
Db 2221 KLFNEIVPLYTLNVNSFVERVQFGKIEHTPTLVHLDISNNLTLDAVRAFAVERQ----- 2276
QY 824 LKTLPLPEYIYGLNVPYFDK--SGKLIIVALDMSDTIG--EG---QKDAYGNPILNDE--- 874
Db 2277 -----LQNFVPGGN--RFDTFVEIDLRAFDDLETIGLEGTVMKCDFLKSLSKASMEADV 2329
QY 875 -----DNEGYH--ALAVATLADYEGLDIKTLNLSKLSQTL-----SIRQVPTAAYHRAGIF 923
Db 2330 TPAYKEKETLHESGLRFATPSTASLSI--TPSDYVYSHVTPENSSLSAIPQTVPAKSGV 2388
QY 924 QAIQNAABAEQLLPKPGTHSEKSSSESANSGKRGSLGNPKTNRGRHSAILPRTGSKGS 983
Db 2389 EFTNASGCDIYELLPKTRFH-----AADLKLYLDNRR-----LRKT----- 2425
QY 984 FVYG--ILGYTSVALLSL 999
Db 2426 --YGSWFLGARSLELLSL 2441

```

RESULT 43

```

Q98QZ9
ID Q98QZ9 PRELIMINARY; PRT; 3216 AA.
AC Q98QZ9;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein MYPU_2110.
GN OrderedLocusNames=MYPU_2110;
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084; DOI=10.1093/nar/29.10.2145;
RA Chandaud I., Heilig R., Ferris S., Barbe V., Samson D., Gallissou F.,
RA Mosser I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis.";

```

```

RL Nucleic Acids Res. 29:2145-2153 (2001).
DR EMBL; AL445563; CAC13384.1; -.
DR PIR; C90538; C90538.
DR MyPUList; MYPU_2110; -.
DR InterPro; IPR007326; Lipoprotein 17.
DR Pfam; PF04200; Lipoprotein 17; 5.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 3216 AA; 359419 MW; A62ED957568DB90C CRC64;

Query Match 4.2%; Score 212.5; DB 2; Length 3216;
Best Local Similarity 19.9%; Pred. No. 1.9;
Matches 218; Conservative 148; Mismatches 380; Indels 347; Gaps 55;

QY 7 TVALTLTT--VSVVTHNQEVPSLVK-----EPILKQTQASSISGADYAESGGSKLKLIN 59
Db 2274 TLTITISSDQTSRTTYQIEGFSVDRNAENIIVAVGYLSEIOASGRAPSGTTVQI- 2332
QY 60 ETSQPVDD-----TVTDLPDSKRTTPE--KIKONLAKGPREQELKAVTENTESEKQITS 111
Db 2333 ----PILDPSVIQNNSTDEFVNKILTNHESLGININKNPQ-----SNNSNLEQSILS 2381
QY 112 GSQI---EQSKESLSLANKTVPESTN-----WEI--CDFITK--GNTLVGLSKSGVEK 156
Db 2382 AEVLTTNEYPAESVKLIQVKSKKGDQDHYSRVWVYVSGFIQRTGADILNIRRAIBAS 2441
QY 157 LSQTD---HLV-----LPSQAADGTQLIQVASFAFTPK-KTAAEYTSRAGE 200
Db 2442 LRENDKAQHIKVKKESREKITKTISTTTVGKNFTTSDPEINNELIKTKVTNVQINDY 2501
QY 201 NGEI--SOLDVDGKEIINEGEVFNYSLLKVKVTIPTGYKHIGQDAFVDN--KNIAEVNLEPE 256
Db 2502 NVEILSVLDISRENIISGNARVKIRSKK-----DNSSEKSEFIVNI--S 2544
QY 257 SLETISDYAFALAKLQIDLPD-----NLKAIAGELAPF---DNQITGLSLPRQL 303
Db 2545 GFETLSEAFDKYIKSSTRFIPSVKEANKDLWDTKATSENFSTPSQNKINDKPDL--- 2600
QY 304 MRLAER-----AFKSNHIIKTIEPRGNSLKVIGEASFQDNDLSQMLPDGLEKIEBS 353
Db 2601 -AVKERFVFSVDTSKVFKFMBLSKTIKTINFGIBSTFQSAQYEDV-----DGFKK--- 2650
QY 354 EAFTGNPDGDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFT 413
Db 2651 ELFDTN---DLFR--VOYEKDGK-----TINKYVKE--KNSHNIDFTK----- 2688
QY 414 YQKNSVTGFSNGLQKVRKNKLEIPKQHNQVTTITEIGDNAPRVNDFQNTLRKYDLEEV 473
Db 2689 -----IKKLEI-----GPNFSEIPKDYFKDAS-----SLTEL 2715
QY 474 KLPSTIRKIGAFAPQSNLKSFEASDDLEIEKEGAFMNNRIETLELKDKLVTIGDAAPHI 533
Db 2716 KINSGVSKIKESAFESAKLTSLPLNSLVEIGNPAFNKNSVLTSL----- 2759
QY 534 NHIVAIVLPESVOEIGRSAPFRONGANNLIF--MGSKVKTIGEMAFPLNRLHLDLSEKQ 591
Db 2760 -----SGLEQTKLSALKEN-----VPEKANDSIKTT-----IWNKKELALKKEPI 2800
QY 592 LTEIPVQAFSDNALKEVLLPASLKTIRBEAFKQKHLKQLEVASALSHIAPNALDNDGDGE 651
Db 2801 LT-----SINK-----IPSED---KKAKKPSQFEQ-----SNFSPAATKQEDSDI 2838
QY 652 QPD-----NKVVVTHNSVALADGEHFIVDPDKLSSTIVDL-----EK 690
Db 2839 HFEITGFSQDDVFGFTIETTYKIKIKTKTDESHPVFDEN---NPSK--KVLDTFTNLEEK 2892
QY 691 ILKLEIGLDY-----STLQTTQTPQPRDMTTAGK--ALLSKNLRQGEKQFLQEAQF 741
Db 2893 VTKIKDASSYFDSKIEETIEKKKOTSSNNTSTENQNKYRLVQWKDQVNGKSGSVFLPQSSV 2952
QY 742 ---FLGRVDL--DKAIAKAEKALVTKKATPKGQLLE--RSINKAVLAYNN----- 784
Db 2953 NAPPVETTTQIGDKTTSLVDFNPPQSGQIRNNLSFENLKPSSAVLVWSNDASLKNQSKIT 3012

```


RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
 RA Richardson P., Rubin E., Rice H.;
 RT "Complete genome sequence of *Bacillus thuringiensis* 97-27.";
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 an amide bond (By similarity).
 DR EMBL; AE017355; AAT61022.1; -;
 DR GO; GO:0009986; C:cell surface; IEA.
 DR GO; GO:0005727; C:extrachromosomal circular DNA; IEA.
 DR InterPro; IPR001434; DUF11.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR Pfam; PF01345; DUF11; 4.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR TIGRFAMs; TIGR01451; B_ant_repeat; 19.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 KW Cell wall; Complete proteome; Hypothetical protein;
 KW Peptidoglycan-anchor.
 SQ SEQUENCE 3471 AA; 386275 MW; C27F7C668311ABA7 CRC64;

Query Match 4.2%; Score 212; DB 2; Length 3471;
 Best Local Similarity 19.8%; Pred. NO. 2.2;
 Matches 215; Conservative 166; Mismatches 362; Indels 342; Gaps 57;

QY 44 GADYAESGSKKLKINETSGPVDDTV-----TDLFSDKRTTPEKIKONLAKGP 91
 DB 1225 GLEYVENSIAKAB---GEAPSPVELTVAGVVKAKYIDITDKERSIVFVKVKEVEVG- 1280

QY 92 RQELKAVTENT-----ESEKQITSGSQLESQKES-LSLNKTV--PSTSNWEICDP-ITK 142
 DB 1281 KEIVNKAIVDDTKHQPIPERIT-----PQKQGIINAAKIVDNPSPKLEEVEYRISF 1335

QY 143 QNTLVGLSKSGVEKLSQTDHL-----VLPQADGTQLIQV-----ASFAPPD 186
 DB 1336 KNT---VENGKLEKVKIEDTTPNGLEYVKGSEKAGDKPAPVKLVKDGKVIAYENITD 1392

QY 187 KKAIAEYTSRAGENETISQLDVGKEIINEGEVENSLLAKKVTIPTGYKH----- 237
 DB 1393 TKERSIVFTVKKEAEI-----GKEIVNQAIVDVDTKDSKPEAKITPLHKOGKIKAKK 1446

QY 238 -----IGQDA-----FVDNKNIAEV-----NLPESETISDYAFAPHLALKQIDL 277
 DB 1447 SVNNETPKLEEVEYRISFKNVTENGKLAEVKIEDTLPGLGYVE----- 1491

QY 278 DNLKAIG-----ELAFDQNIQTKLSLPRQMLRAEA--FKSNHIKTIIFRGNLSKVI 329
 DB 1492 NSLKAEGAGNPVELKMGKVLAKY--ETTDIERSITPK---VKVSEVKVGEKIV 1545

QY 330 GEASQD-----NDLSQL--MLPGLKIESEAFGNPG---DDHYNNRVVLTKSGKN 378
 DB 1546 NKAIVDDTKHDPINPKASITPQYKDG--KIEAEKVNNPSPKLEEVEYRI-----SPKN 1598

QY 379 -----PSG-----LATENTVYVNPDKSLWOES-----PEIDVTKWL 409
 DB 1599 TVEHGKLAEVIEDDLPNGLEYVKSQLEAGSKPNPVELKPKENGKVLAKYFEITDK--E 1656

QY 410 EDFTYQ---KNSVTG--FSNGKL-----QKVRNKN--LEIPKHQNGVTITEI 450
 DB 1657 RSITPKVKVKNVSDTIINQAIVSDTKHPPETPKAEIIPQHKDGKLEAKVNNL-LPKL 1715

QY 451 GDNAFRNVDPQNTLRKYDLBEVKLPSTIRKIGAFQFOSNLKSPSEASDDLEEIKEGAFM 510
 DB 1716 GEEVEYRISFPK--TLENGKLAEVKIEDTLP--GLEYYVENSILKAEAGTDSVELK---FE 1769

QY 511 NNRI-----ETLELKDVLVTIGDAAPHINHIYAIVLPSVQVIBGRSAFRQNGANNLI 565
 DB 1770 NGKILAKYPEITDKERSIT-----PKVKVKEVTKIGKI-----VNKAIDDD 1812

QY 566 SKVK-----TLGEMAFNLRLBHLDLSEQKQITETIPV 597
 DB 1813 TKKEPTTAETIPQHKDGKVEAEKTVNNPSPKLEE--IEYRISFKNVTENGKLAEVKI 1870

QY 598 QAFSDNALKEVLLPASLTIREAEFPKNHLKQLEVASALSALSHAFNALDNDGDGEQFDN 657

DB 1871 EDTLPNGLEYV-----KDSLQAGSKPNPV-ELKVDGKVIKYPALITDI--ERSI 1919
 QY 658 VVKTHNSYALADGEHFIVDPDKLSSIVDLKELKLEIGLDYSTLRQTQTQTPRDMTTA 717
 DB 1920 VFKAQV-----EDFKVGBGIQNVVVDTK-----DPKTRVTVTPPKD--- 1960

QY 718 GKALLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKQGLLERSINK 777
 DB 1961 GK--LKAKEKPVNKKPKLGEVEY---RINF-----KNTVENGKLEVEVKVED 2002

QY 778 AV---LANNNSAIKKNVKKLEKLELDLTGLVEGCKPLAQATMVQGVV---LKTPLPLP 831
 DB 2003 EIPAGLSYVNSLQAGSKSPVELKFKENGKVMKAPVITDKERSIVFVKAKVBEAEIG 2062

QY 832 EYITGLNVYFD-----KSGKLIYALDMSDTTIGBQKQDAYGNPILNVDED 875
 DB 2063 KEIVNKAIVDDTTHPEKPYVEITPQYKDGKIV-----AEKVANNHKPKLGEV--- 2111

QY 876 NEGYHALAVATLADYEGLDIKTLNLSKLSQTSITRQVPTAAYHRAGIFQAIQNAAEAEQ 935
 DB 2112 -----BYRIRFKNTVESGKLAEVNIKDTLPKGLGYVEG-----SITAEG 2150

QY 936 LLPKP 940
 DB 2151 SKPKP 2155

RESULT 46
 Q6BY65 PRELIMINARY; PRT; 1133 AA.

AC Q6BY65;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Debaryomyces hansenii chromosome A of strain CBS767 of Debaryomyces
 DE hansenii.
 GN ORFNames=DEHA0A12507g;
 OS Debaryomyces hansenii CBS767.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
 OC NCBI_TaxID=284592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS767;
 RG Genolevures;
 RA Lafontaine I., de Montigny J., March C., Neuvéglise C., Talia E.,
 RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykaeten C.,
 RA Boisarame A., Boyer J., Catolico L., Confaniolieri F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hanttrave F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Kozul R., Lemaire M., Leur J., Ma L., Muller H.,
 RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenn S., Pottier S., Richard G.F., Straub M.L., Suleau A.,
 RA Swennen D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierly A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RA Winkler P., Souciet J.L.;
 RT "Genome evolution in yeasts.";
 RL Nature 430:33-44 (2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS767;
 RA Genoscope;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; CR382133; CAG84829.1; -;
 DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
 DR GO; GO:0006259; P:DNA metabolism; IEA.
 DR GO; GO:0000160; P:two-component signal transduction system (p. .); IEA.
 DR InterPro; IPR000237; GRIP.
 DR InterPro; IPR008207; Hpt.

```
DR InterPro: IPR007696; MutS_III.
DR Pfam: PF01495; GRIP; 1.
DR PROSITE: PS0913; GRIP; 1.
SQ SEQUENCE 1133 AA; 130147 MW; 799476682BEE9EC2 CRC64;

Query Match
Best Local Similarity 4.2%; Score 211.5; DB 2; Length 1133;
Matches 233; Conservative 156; Mismatches 383; Indels 337; Gaps 54;

QY 30 EPIKQTOASSISGA---DYAESGSKSLKINETSGPVDVTDLFSDKRTTPKIKD 85
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 EKVLQENTPISISDAGLIVEYLNG-----VNEKTAMLEEMRKVAKENSLSNEKLD 183
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 NLAKGPREQELKAVTENTSEKQITSGSQLEQSKESLSLNKTVTPSTSNWEICDFITKNT 145
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 -----VEARKL-----EALAKDKSEGEVTKDSISSQESNDELKLN 222
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 LVGLSKSGVEKLSQTDHLVLPQAADGTQLIQVAFPTDKKTAETATYTSRAGENGEIS 205
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 TISDLKGLKLEK-----KQAEKDT-----VEEKGENKKEE 252
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 206 QLD-----VDG--KEIINEGEVFNLYLLKKVTIPTGYKHIGOD--AFVDNKNIAEV 252
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 253 KYDRDTKQSEVDGKDKIESYKKEIESVKLVKSESSSRNKKDELGAALERSN----- 308
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 253 NLPSELETISDYAFAHALKQIDLDPNLKAIGELAFPDNQITG--KLSPRLQMLRAERAF 311
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 -----SLEKDKLNDMSKNDNETIGLTKLEEKYKKQLAELVD 346
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 312 KSNHNT--TEPRGNSLKVIGEASFPQNDLSQMLPD-----GLEKIESEATGNP-GGD 363
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 VNSALETEIENKKELEK-----NFNDISGTMQNDLGNANKSIENLKSQAELNERASD 399
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 364 HYN-----NRVLWTKSGKGNPSGLATENTYVNPDKSLWOESPEIDYTK----- 406
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 400 LNLQDLKRNKIKELOQNSQPIEAT--TPD-----TPSSNNKNNKNNKGGKKN 451
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 407 -WLEEDFTYQKNSVTGFSNGKQKVRKNRKNLEIPKOHNGV--TITEIGNAFRNVDFQNK 463
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 452 STTDNQIIVKSESEQLSAK--PEDIQNYDELKKNERIQLCKEIDSSKQEISDELKR 509
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 464 TLKRYD--LEE-----VKLPSTIR-KIGAPAFOSNNLKGF-----EASDLEBEIK-- 505
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 510 TKSKYDKLVEEHDLLITNHNKVKSLDKLKIETDVHLRLDLLTKTGDDLVQARDEIKELKAK 569
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 506 -EGAFMNN-----RIETLELKDKLVTIGDAAFINHIYAIVLPESVQEIERSAFRQNGANN 560
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 570 SKSVKLNDCEUKKELTEMQAKPKELNESKEDIE-----KSKQELKS--IEDNKSNN 620
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 561 LIFMGSVKVTL---GEMAFSLNRLEHL--DLSEQKQLT-----EIPVQAFSDNALKE 607
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 621 -----SSNEKLLSTVRGEKQKLENSFRELTKDLEASKEKNSDLEKELEKLSIEDNDLSQ 676
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 608 VLLPA-----SLKTRIRSEA-----FKKNHLK--OLEVA 633
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 677 SLNKADHDALNEKHSALKSKFBEETNQFEIAKQEKDILNRIKELSEFKSNDTSLKLETA 736
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 634 S--ALSHI-----APNALDDNDGDEQFDNKVVVYKTHNSYA 667
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 737 SLKTSISHQDLIENFKKIDELNKLNEELSGSISKLNASNDELQNSKDLVSEKRN---T 793
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 668 LADGEHFI VDPD-----KLSSTIVDLKILKILEGLDYSTLR-----OTTOTQ 710
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 794 LVTKQELSFEKNNLSNNELLKQSEKQKLSLETMTWKDKHDSLVKDKDTSSSSSIQTFRQQ 853
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 711 FRDMTQAG--ALLSKSNLRQ--GEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKATKN 767
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 854 YEELSMSKESLRLIDNDELNESRTMLQERT--RESSTIRRLVDAEQLKLNHDPK 911
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 768 GOLLERSINKAVLAYNNSAIKANVKRLEKLELDLLTGLVEGKGPLAQMVGCVILLKTP 827
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 912 CELTKISKEKSEVENNCOAL-----IKKQREIDEFKSIDTN-----YLLKIQ 954
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 47
O07910
ID O07910 PRELIMINARY; PRT; 1247 AA.
AC O07910;
DT 01-JUL-1997 (TremBLrel. 04, Created)
DT 01-JUL-1997 (TremBLrel. 04, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Caga.
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43526; and ATCC43526;
RA Maeda S., Kanai F., Ogura K., Yoshida H., Ikenoue T., Kawabe T.,
RL Shiratori Y., Omata M.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001357; AAB58747.1; -
DR EMBL; AB003397; BAA19853.1; -
DR HSSP; O9AGB1; IFIM.
DR GO; GO:0019534; F:toxin transporter activity; IEA.
DR InterPro; IPR005169; Caga.
DR InterPro; IPR004355; IVSec_caga.
DR Pfam; PF03507; Caga; 3.
DR PRINTS; PR01553; TYPE4SSCAGA.
SQ SEQUENCE 1247 AA; 138819 MW; 172C76AE9F9BD29E CRC64;

Query Match
Best Local Similarity 4.2%; Score 211.5; DB 2; Length 1247;
Matches 226; Conservative 146; Mismatches 389; Indels 329; Gaps 54;

QY 27 LVKEPILKQTOASSISGADYAESGSKSLKINETSGPVDVTDLFSDKRTTPKIKDN 86
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1 MTNETIAQQPQTEAFNPQQFINNLQVAFK-----VDNAVASYPDQKPIVDK---- 49
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 87 LAKGPREQELKAVTENTSEKQITSGSQLEQSKESLSLNKTV--PSTSNWEICDFITKGN 144
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 50 -----NDRNRQAFDG--ISQLEBEYS--NKAINKPTKKNQYPSDFINKSN 91
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 TL-----VGLSKSGVEKLSQTDHLVLPQAADGTQLIQVASF-----TP 185
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 DLINKNDLIDIGSSIKSFQKFGCTQRYRFTSVSHQNDPSKINTSIRFMENIIQPPTP 151
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 DKKTATAIAYTRAGEN--GEI--SOLDVDGKEINEGEVFNLSYLLKKVTIPTGYKHIGOD 241
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 DKKER--AEFVKSASQFAGIIGNQIRTDQKFM-----GVDFEFLKERQEAENKGEPTGD 206
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 242 -----AFVDNKNIAEVNLPESLETISDYAFAHALKQIDLDPNLKAIGELAFPDNQITG 295
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 WLDIPLSFVFNKEOSS-----DVKEAINQEPVPHVQ-----PD-----IATTTTHIQG 249
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 296 KLSLPROLMRLAERA-----FKSNHKTIEFRG-----NSLKVIGEASFQNDLSQ 341
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 LPPESRDL--LDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLI-----HNNALSS 301
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 LML--PDGLEKIE--SEAFNGNP-----DDHYNRNVVLWTKSGKNSPGLAT 384
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 302 VLMGSHNGIEPEKVSLLFAGNGGFGAKHDWNATVGYKQKQGDVATLINVHMKGSGGLVI 361
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

QY 395 ---ENTYVNPDKSLWQESPEIDYTKWLEBDFYQKNSVTGFSNKGLOKVKRNK--NLBIP 439
Db 362 AGGEGGINP-----SFYLYKEDQLTGSQRALSQBEIRKIDFMEFL 403
QY 440 KQHNGV--TITEIGNAFRN--VDFQ-----NKTLRKYDLEEVKLPSTIRKIG- 483
Db 404 AQNNAKLNLSKEKEKQNEIEDFQDSKAYLDALGNDRIAFAVSKOPKHPALITEFGK 463
QY 484 ---AFAFOSNNLKSFEASDDLEIEKEGAFMNNRIETLELKOKLVITGDAAFHNNIYAI 540
Db 464 GDLSTYLVKMGKKQTEALD-----REKNTVLQGNLKH----- 495
QY 541 LPESVQETGRSAPRONGANNLIFMGSKVKTLCGEMAFNLNRLHLDLSQKQLTEIPVQAP 600
Db 496 --DSVMFYNSFKYTNASK-----SPDKGVG-----VTNGVSHLDAGFSK-----VAVF 538
QY 601 SDNALKEVLLPASLKTIREEAFAKQKHLKQLEVASALSHIAPNALDNDGDQFQNKVVV- 659
Db 539 N-----LP-DLNNLAITSFVRNLENKLVTEGLSLOEANKLIK-----DFLSSNKLUG 586
QY 660 KTHNSYALADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSTLRQTTQTQFRDMTTAGK 719
Db 587 KALNFKAVADAKN--TGDYDEVKKAQKDLKSLRKREHLE-----K 626
QY 720 ALLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAK-----ALVTKATKNGQLLERSI 775
Db 627 EYEKLETSGNKNKWEAKAQ-----ANGQDKIFALINKEANRDARAAYSQ 674
QY 776 N-KAVLAYNNSAIKKAAN--VVRLEKELDLTLGLVEGKGLAQATWVGYYLLKTPLELPE 832
Db 675 NLKGTRELSDKLEKINTDLKDFSKFD-----E 703
QY 833 YYIGLVNVPDKSGLIYALDMSDTTIGEGQKDAYGNP--ILNVDEDEGYH----- 880
Db 704 FKNGKNDKFSKAEETLKALKGS-----VKDLGINPEWISKVENLNAALNBPNGKNKDF 757
QY 881 ALAVATLADYEGLDITKILNSKLS--QLTSIROVPTAAVHRAGIFQAIQNAABEQLLPK 939
Db 758 SKVTOAKSLENSVKDVIINQKITDKVDNLAQVSMKATGG--FGRVEQALADL----- 810
QY 940 POTHSEKSSSSANSKORGLOSNPKTNRGHSAILP--RTGSKGSFV--YGLVYTSVAL 996
Db 811 -----KNFSKEQLAQO-----TQKNSFNVGKKESEIYQSVKNGVGTLVNGLSGIEATAL 861
QY 997 LSLTAIKKK 1006
Db 862 AKNPSDIKKE 871

RESULT 40
Q9AHK6
ID Q9AHK6
AC Q9AHK6;
DT 01-JUN-2001 (TreeBLrel. 17, Created)
DT 01-JUN-2001 (TreeBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TreeBLrel. 26, Last annotation update)
DE LMP1.
GN Name=lmp1;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB19;
RA Dunn J.J.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF305609; AK18801.1; -.
DR HSRP; P50542; IFCH.
DR InterPro; IPR002713; FF.
DR InterPro; IPR008940; Prenyl trans.
DR InterPro; IPR006597; Sel like.
DR InterPro; IPR001440; TPR.

```

DR InterPro; IPR008941; TPR-like.

DR Pfam; PF01846; FF; 7.

DR Pfam; PF00515; TPR_1; 4.

DR SMART; SM00671; SEL1; 4.

DR SMART; SM00028; TPR; 9.

DR PROSITE; PS50005; TPR; 6.

DR PROSITE; PS50293; TPR_REGION; 1.

KW Repeat; TPR repeat.

SQ SEQUENCE 1173 AA; 134287 MW; F0713ADF2E2311DC CRC64;

Query Match

Best Local Similarity 4.1%; Score 210; DB 2; Length 1173;

Matches 224; Conservative 177; Mismatches 393; Indels 404; Gaps 57;

QY 8 VALFTTVSVVTHNQEVFSLVKPEILKQTOQASSISGADYABSSGSKLKINETSGPV-- 65

Db 75 IAMLNTLAEIANSPPES-----KDLQDSANQILDKIKGQDNTKTNNVNFDAFN 127

QY 66 -----DDTVTDLPD-----KRTTPEKIKONLAKGPREQE----- 95

Db 128 NRYIKDSTITENYSRNDVDGIEDIEDISEFKKSKIPKIPN--TNPKEBQIIQSNNPK 185

QY 96 -----LKAVENTESEKQITSGSOLEQESLSLN----- 125

Db 186 LSVNDQKNLNLKLNKLSGKNSNSENLANSQKIENDKQNTLSKEKNSENILKTPDNS 245

QY 126 -----KTVPSTSN-----WEICDFITKGNL 146

Db 246 KYSNNNTTSLKIPSNQSESELSPSQTIIGKIYRPYSYLIKELYELDDINTGRVT 305

QY 147 VG-----LSKSGV-EKLSQTDHLVLPQAADGTQLI-----QVASFATPDOKKT 189

Db 306 LGRNLKELIKGLSNKPKQVNLNLSKNSKNSKNSKNSKNSKNSKNSKNSKNSKNSK 364

QY 190 AIAEYTSRAGENGISOLDVDGKEIINEGEVFN--YLLKKVTIPTGYKHIGQDAFVD-- 245

Db 365 -----EIFQDKGDKKQHPGDLKSKVHSIKPIDLENTKSR--QQAIKOLNE 409

QY 246 -----NKNIAEVNLPESLETISDYAFAPHLAKQIDLPNLK-----AIGELAPF- 289

Db 410 FLKNNPDAQASKTLAQANKIQHLEDLKSRLVH--SIKPIDL-ENTKSRQQAIKDLNEFL 465

QY 290 -----DNQITCKLSLPQLMLAERAPKSNHIKTIEP-----RGNLSKVIGEASPDND- 338

Db 466 KNNPDAQASKTLAQANKIQHLEDLKSRYISIRPIDLENTKSRQQAIKOLNE--FLKNNP 523

QY 339 -----LSQLMLPDGLEKIESEAFNCPGDDHNNRVLVLTWSGK-----N 378

Db 524 NDAQASKTLAQANKIQHLEDLKSRYISIRPIDL-LEN-----TKSRQQAIKDLNEFLKNN 576

QY 379 PSGLATENTYVNPDK-----SLWQESP-EIDYTKWLEEDFT-----YQKNSVTGFS 423

Db 577 PNDQAASKTLAQANKIQHLEDLKSRYISIRPIDLENTKSRQQAIKDLNEFLKNNPDAQA 636

QY 424 NKGLQKVKENKNIKLEIPKQHNQVTTIEICDNAFRNVDFQNK--TLRKVDLEEVK-LPSTIR 480

Db 637 SKTLAQANKIQHLE-----DLKSKVYSIRPIDLENTKSRQQAIAK 675

QY 481 KIGAFAFOSNNLKSFEASDDLEIEKEGAFMNNRIETLE-LKDKLVITGDAAFHNNIYAI 539

Db 676 DLNEFL--KNNPNDQAASKTLAQ-----NKIQHLEDLKS-----VHSIRPI 716

QY 540 VL-----PESVQIGRSAPRONGANNLIFMGSKVKTLCGEMAFNLNRLHLDLSQKQLT 593

Db 717 DLENTKSRQQAIKDL--NEFLKNNPND-----AQASKTLAQ-----ANKIQHLEDLKS 766

QY 594 EIPQAFSDNALKEVLLPASLKTIREEAFAKQKHLKQLEVASALSHIAPNALDNDGD-- 650

Db 767 IRPIDLENTKSRQO-----AIKOLNE--FLKNNPNDQAASKTLAQ--AY-----ENNGDLK 814

QY 651 -EQFDNKKVVKVT-----HHN-----SYALADGEHFI-----VDPKLS--TIV 686

Db 815 AENAYEKIKLNTQEDHYKLGIRPKLKKYEHSEIESPDQTIKLDPKHKKALHNKGIALM 874


```

QY 687 DLKILKLEGLDYSTLROTTOTQFRDMMTAGKALLSKSNLRQGEKQKFLQBAQ----- 740
D 785 MLNKNKKAIESPE-----KAIQIDKNYGTAYYQKGIABKNGDMQ 915
QY 741 ----PFLGRVDDLDKAIKAKSKALVTKATKNGQLLERSIN-----KAVLAYNNSAIK 788
D 916 AFASFKNAYLNDKNPNYALKAGIVSNLGNFKQSEYLNFFNANAKPNEIAIYNLSIAK 975
QY 789 KANVKLELEKLDLLTGLVE-----GKGLAQATWQGVVLLKTPPLP 831
D 976 PEN-NKLESLETINKAIDLNEPEKSEYLYLKASINLKNENYQVAILSYLVIEKAP-ENT 1033
QY 832 EYVIGLVNVPDYSKGLIYALD-MSDTIGEGQDAYGN-----PIL 870
D 1034 SAYINLAKAYEGSGNKSQAISTLEKIINKNNKLNALNGLYKKEKNYQKATEIFEKAI 1093
QY 871 NVDENEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTTAHYRAGIFQAIQ 928
D 1094 N--SDIEAKYNLATTILIEINDNTRAKDLR-----EYTKLPNNPEALHALGIIENEN 1145

RESULT 49
Q9BJY0 PRELIMINARY; PRT; 2752 AA.
AC Q9BJY0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 235 kDa rhostry protein (Fragment).
GN Name=Iifa.1;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21273138; PubMed=11378199; DOI=10.1016/S0166-6851(01)00253-5;
RA Khan S.M., Jarra W., Bayele H., Preiser P.R.;
RT "Distribution and characterisation of the 235 kDa rhostry multigene
family within the genomes of virulent and avirulent lines of
Plasmodium yoelii."
RL Mol. Biochem. Parasitol. 114:197-208 (2001).
DR EMBL; AF323442; AAK15625.1; -
DR InterPro; IPR006499; ReticulocyteBP.
FT TIGRFAMs; TIGR01612; 235kDa-fam; 1.
FT NON_TER 1
FT NON_TER 2752
FT NON_TER 2752
SQ SEQUENCE 2752 AA; 322608 MW; E834E3753168AF76 CRC64;

```

```

Query Match 4.1%; Score 210; DB 2; Length 2752;
Best Local Similarity 20.5%; Pred. No. 2;
Matches 201; Conservative 162; Mismatches 355; Indels 262; Gaps 47;

QY 7 TVALTLTVSVVTH-NQEVFSVLGPIKQTOASSI-----SGADYA 48
D 479 TISLILPXMDFKTNLNSVMVQLKNGIKLKFSSQIKNKNLVSTVPERKEGFLSSIELA 538
QY 49 ESSGSKLKI-----NETSGPVDVTVDLSDKRTTPPEKIKDMLAKGPRQELKAYTE 101
D 539 KSWKEKEEIIKKIENEBETVQLODKIRELP-----KKYLDLAEEKYIEDLKLEIN 591
QY 102 N-----TESEKQITSGSQLESKES-----LSLNKTVRSTSNWEICDFITKNTLVGLSK 151
D 592 NKIKDITEKIKYKXKTVLDLKEVEKDVVIDELAKQPP-----YQTEYEKKNTYIDYIK 647
QY 152 SGVEKLSQTDHLVLPSSQADGTLQIVASFAFTPKDKTAIAEYTSRAGENGISQLVDVG 211
D 648 SDIKQIYVDDIDLLYNEMS-----SVLQENTINNENK--TELETUK 687
QY 212 KEIINEGEVFNYSLLKKVTIPTGYKHIGODAFVDNKNIAEVN---LPSLETISDYAFAH 268
D 688 SKIDN---YNNK-----IKHLETEAVEKRLNNIETNNKNSLGSITLBIKKYIYGE 733

```

```

QY 269 LALKQID--LPDNLKAIAGELAFFDNQIT---GKLSLPRQLM----- 304
D 734 ID-KELNKTLDNPKNQEQELSNKINDYTKENDKLSVYKSKILLEIRYNYNDQINIDNTKGE 792
QY 305 RLAEAFKSN-HIKTIEFRGNSL-KVIGEASFQDNDLSQMLPDGLEKIE---SEAPT 357
D 793 EAKENYKSNBHMKTPTTNEYEISEAINEVKRMKNEL-----LSKVYKVPFDNIYN 844
QY 358 GNPGDHYNRVWLWTKSGKN---PSGLATENTYVNPDKSLWQESPE-----IDYTK 406
D 845 GNVDESH--NKFIDLTNKKNEVSDEKLSKHETKENGSKSLINETKKSIDIDYQINITLK 902
QY 407 WLEEDFTY---QKNSVTG-FSNKGLQKVRKNQLEIPKOHNGVTTITEGDNAFRNVDFON 462
D 903 KVDEYIKVCKRTKESITNFFSKQTTLKDKLNQINTIKKIDPI-----EKPYKD-QPEN 955
QY 463 KTLRYDLEEVKLPSTIRKIGAFAPQSNLKSFEASDDLE-----EIKEG 507
D 956 RLINKIS-----ELDKKPKDASLXDHESNNNGLMFEYNNLKANLGNKKEKTLHYHEFDEK 1011
QY 508 AFMNRRIETLEKOKLVT-----IGDAAFHINHIYALVLPESVQIEGRSAPRONGANNLI 562
D 1012 A-VNNIITIEDINKNISIEIYTSYINIEEDAIKKNIKLLNDQVVKVKNVNTN 1070
QY 563 FMGSKVTLGEMAFSLNRLHLDLSEKQKLTPIVQAFSDNALKEVLLPASLKTIREAP 622
D 1071 LSGIKEKL-----KRYNFDLFGKBNKIYI-----NEIKINDINTLDQKXD 1112
QY 623 KKNHLKQLEVASALSHIAFNALDDNDGDEQFNQKVVVTHNSYALAGEHIVDPDKLS 682
D 1113 ----KSTLTIKTKTSESHIGEIKGTQDKLEK-----ADINTYXEDPKSIE 1156
QY 683 STIVDLK-----ILKIEGLDYSTLROTTOTQFRDMMTAGKALLSKSNLRQGEKQK 735
D 1157 KKIENVVKKDKKNYKEINKL-----LNEISEIE-KDKTSLEE--LKNINLSYGRSL-- 1207
QY 736 LQEAQFFILGRVLDLKAIAEAKALVTKATKNGQLLERSINKAVILAY--NNSAIKKNVYK 793
D 1208 --GNIFLEQIDEEK-----KKA-----ERTI-KAMEAYIEDLDNKKKS-D 1244
QY 794 RLEKELDLTLGLVEGKGLAQATWQGVVLLKTPPLPEYVIGLVNVPDYSKGLIYAL-- 851
D 1245 EIEKDKMIKMDINEEMKALNISNDDRNYHTKS-----KDHKKGISDIHDKSKSIQNF 1300
QY 852 -----DMSDTIGEGQK 862
D 1301 ESDINNIKNEQENVSESRK 1320

RESULT 50
Q97K41 PRELIMINARY; PRT; 1227 AA.
ID Q97K41
AC Q97K41;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Uncharacterized protein, probably surface-located.
GN OrderedlocusNames=CAC1080;
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RX DOI=10.1128/JB.183.16.4823-4838.2001;
RA Noelling J., Breston G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabatne F., Doucetce-Stamm L.A., Soucaille P.,
RA Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum."

```

```

RL J. Bacteriol. 183:4823-4838 (2001).
DR EMBL: AE007624; RAK79054.1; -.
DR PIR: C97033; C97033.
KW Complete proteome.
SQ SEQUENCE 1227 AA; 134030 MW; 66EBE3399AF6FB38 CRC64;

Query Match
Best Local Similarity 4.1%; Score 209.5; DB 2; Length 1227;
Matches 237; Conservative 207; Mismatches 402; Indels 463; Gaps 68;

QY 1 MKKHLKTVLTLTVSVVTHNOEVSFLVKPEPLKQTAQSSISGADYAESGSKLKINE 60
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 TSGPVDVDTLDFSDKRTPEKIKONLAKPREQELKAVNTESSEKQITGSQLBQSK 120
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 50 KGKEV-----QEKEVNVKNNDNSKSVSSNE-NQVSNKENSNPVSSSEIQSINK 100
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 SLSL-----NKTVPSTSNWEICDFITKGT-----LVGLSKSGVEKLSQTDHLVP 166
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 101 NVNLQVQNNKSVLAASN--VDDEVKKNNGVNTSDYIAIGETVKVPTDLDIINKAIVD 158
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 SOAADGTOLIQVAFPTPKKTAIAEYTSRAGEN-GBISQIDVDGKEIINEGEV--FNS 223
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 ABSAEGTDLSEVEIYDIVSQTAEAIQEAFRINDGVANVSDYTLGATFVNDANLDSYNK 218
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 YLL-KKVTIPTGYKHI-----GQDAFVD-----NQIAEV 252
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 YFYKRYATVTKFQDVATKTSNALKNINNGQGETDYTALEVSGVTQPYLDLVNKNIVKE 278
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 253 NLPE-----SLETSIDYAFHLALKQIDLPDLNKAIGELAFPDNQITKGLSLPRLMLAE 308
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 279 KQNGRDLTITEISAASTIARINTAL--DNMDA--GVATLEDYQAIQANNVP-QL----- 330
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 RAFKSNHIKTIEFRGNSLKV---IGEAQFQNDLSQLM-----LPDGL-----EKTESRAF 356
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 331 -----HVADV-----NSLAMDQWGDVSEAIDGINTIYNNINSNGVGTEDDYINSIAV 380
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 357 TQNGDDHYN--NRVVLWTMSKN-----PSGLATENTYV 389
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 381 DSNENIDYDILNANIIEKTKAGQDLTIPEVANVVKVTKLLDFYNHAAAGQTLQDYK 440
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 390 NPKSLMQESPRI-----SNVSGDGSVDYITIGIDGVTEINI-EFVNERIKESGIT 584
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 517 LELKDKLVITIGDAAPHNIIYALVP--ESVQEIGRSAPRQNGANNLIFMGSKVTLGEM 574
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 585 ITIENIKVI-BPIVOLSEVVRIVTGVTGVYKTLGINNVNDNNIYINAEILKQKVD 643
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 575 AF-----LSNRLHLD-----LSEKQL-----TEPVOAFSDN 603
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 644 KIQDIQTRVDNTINNIDVINKIGADAVLSDFYFNIGITDVYQDILDYVNDADLKIQNYKD- 702
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 604 ALKEVLLPASLKTIREAPKQKHLEVASALSHAFNAL-----DDNQG-----DSQF 653
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 703 -VDDIIEVEAKISSVEA-----LMRINIGEAFTD-DPKALGLTIDNDGLLYATYTDLQN 755
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 654 DN-----KVVVTH-----HNSVALAD-----GEH 673
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 756 KNYKTADEIARVQAIQEIYRALMQINLGKATTADYNTLGIIDVNDSTILTYVNDALQKN 815
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 674 FT-VDPDKLSSTIVLEKILKIEGL-----DYSTLRQTTQTFQFRDMMTAGKAL 721
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 816 YVNVDEVK-----AIEKNIQIYNALLKIDSGSATIDDYRTIGITTVIDYN----- 861
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 51
RBP2 PLAVB
ID RBP2 PLAVB STANDARD; PRT; 2867 AA.
AC Q00759; QSN2M3;
DT 01-APR-1993 (Rel. 25, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Reticulocyte binding protein 2 precursor (PVRBP-2).
GN Name=RBP-2; Synonym=RBP2;
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.
RP MEDLINE=2029192; PubMed=10838229; DOI=10.1016/S0166-6851(00)00219-X;
RA Galinski M.R., Xu M., Barnwell J.W.;
RT "Plasmodium vivax reticulocyte binding protein-2 (PvRBP-2) shares
RT structural features with PvRBP-1 and the Plasmodium yoelii 235 kDa
RT rhoptry protein family.";
RL Mol. Biochem. Parasitol. 108:257-262 (2000).
RN [2]
SEQUENCE OF 1189-2439 FROM N.A.
RP MEDLINE=92315338; PubMed=1617731; DOI=10.1016/0092-8674(92)90642-P;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites.";
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
CC human reticulocyte cells.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/ebis/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: AF184623; AAF76525.1; -.
DR HSP; P03069; ICGM.
DR InterPro; IPR003900; KID repeat.
KW Malaria; Receptor; Repeat; Signal; Transmembrane.
FT SIGNAL 1 21 Potential.
FT CHAIN 22 2867 Reticulocyte binding protein 2.
FT DOMAIN 22 2805 Extracellular (Potential).
FT TRANSMEM 2806 2826 Potential.

```

FT DOMAIN 2827 2867 Cytoplasmic (Potential).
FT DOMAIN 44 133 Asn-rich.
FT DOMAIN 560 758 Lys-rich.
FT DOMAIN 1112 1285 Lys-rich.
FT DOMAIN 2758 2785 7 X 4 AA tandem repeats of H-D-D-T.
FT REPEAT 2758 2761 1.
FT REPEAT 2762 2765 2.
FT REPEAT 2766 2769 3.
FT REPEAT 2770 2773 4.
FT REPEAT 2774 2777 5.
FT REPEAT 2778 2781 6.
FT REPEAT 2782 2785 7.
SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFFD3 CRC64;

Query Match 4.1%; Score 209.5; DB 1; Length 2867;
Best Local Similarity 21.3%; Pred. No. 2.2;
Matches 244; Conservative 176; Mismatches 350; Indels 377; Gaps 66;

QY 23 EVPSLVKEPILKQTOASSISGAD-----YAE---SSGSKLKLINETSFPDVTVDLPS 74
DB 1265 EKIELYKKEIDEIKQNTNEVKGQDTSNFYVTEQYNSATQSKAKIEQ-----FINIAT 1316
QY 75 DKRTTEPKIKDNLAGPRQELKAVNTENTSEKQIT---SGSOLEQSKESLSLNTKTPVST 131
DB 1317 TKRGTSQDTSQD-----INELSIKEEVEHKNLQLVKQESNMEEMRKQILSKOLLILN 1369
QY 132 SNWEICDFTTKG-NTLVGLSKSGVEKLSOTDHLVPSQAADGTQLIQVASFAFTPKKTA 190
DB 1370 NSBTIAKELISNTQNALGFRENKTKLTKIDEL-LQRVAA-----MIEEAK-----AKNN 1419
QY 191 IABYTSRAGENGESIQDVGDKBEIINEGEVFNISYLLK-----KVPTGYKHIGQD--A 242
DB 1420 IDIALEDAQIDTEVSKIEQINREIMNKDBEIKSYLSEIKYDKCKTEISNSKRGDKTE 1479
QY 243 FVD-----NK-NIAENVLPESLETISDPAFAHLAKQI-DLPDNLKAIGEL----- 286
DB 1480 FLEKFKPNEESNKNVINEIN--ENIRNSEQ-----LKQIEDAEKQASTKVELFHKH 1531
QY 287 -----AFPDNQITG-----KLSLPROLMRLAER-----AFKSNHKTJE--- 320
DB 1532 EYTIISNIFKESEILGVETSKQKINKAEDIMKIEIRHNEISIQTVKGFGQENLKNLEPN 1591
QY 321 -----FRGNSLKVIGEASFQD--NDLSQLM-LPDGLEKIES-----BAFTG 358
DB 1592 YDVAEDBLNDKSTNAKVLITNLESVGNLSSEITWIKQGEKIYSKAKIMQIKATSE 1651
QY 359 NPD-----DDHYN-----NRVLWTKSGKPKPSGLATENTVVPDQSLWQBS 399
DB 1652 NTAETKLEKVDQDSNVYVNLQITERNLIVTEKRNLN-----GIDSTITNIEGAL-KES 1706
QY 400 P---EIDYTKWLEE-----DFTYQK-NSVTG-----FSNKGLOKVRKNKL----- 436
DB 1707 KGNYEIGFLEKLEBEIGNRKLKVDITKKSINSTVGNPSSLFNNFDNLQDNFNKNINDYEN 1766
QY 437 ---EIPKQNGVITIBIGNAFNV-----DFQK-KTLR-----KYDL-----EVRKLP 476
DB 1767 KWGEIYNEPEG-SLNKISEN-LRNASENTSDYNSAKTLRLAEQKERVNLLNKEEANKYL 1824
QY 477 STTRKIGAFAPQSNLKSFEASDDLEBEIKGAFMNNRIETLEKOKLVIT-----GDAAPH 532
DB 1825 RDVKKVESFRF-----IFNMKESLOKINE-----MIKKEQLTVNEGHGVNKL 1867
QY 533 INHIYAVLPESVQIEGRSAFRONGANNLI---FMGSKVKTIGEMAFSLNRLHLDLSQ 589
DB 1868 VENIKELVDENNLSDILKQATGKQBEIQKITHSTLKNKAKTI-----LGHVDTSAK 1918
QY 590 KQLTEIPVQAFQSNALKEVLLPASPRTIREAFK-KNHL-----KOLEVASALS 637
DB 1919 -----YVGIKTPELATELLGDAKLTAQELKPEKNNVVLETFENMKNVDELVDVHKNTQ 1974
QY 638 ---HIAFNALDDND-----GDE-----QPDNKV-VVKTHNSYALA 669
DB 1975 DAYKVALEILAHSDIEDTQKQDSKLIEMGNQIYLVKVLINQYKNKISSIKSEAEVSVK 2034

RESULT 52

Q6CPF6 ID Q6CPF6 PRELIMINARY; PRT; 1755 AA.
AC Q6CPF6;
DT 25-OCT-2004 (TremBLrel. 28, Created)
DT 25-OCT-2004 (TremBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TremBLrel. 28, Last annotation update)
DE Kluyveromyces lactis strain NRRL Y-1140 chromosome E of strain NRRL Y-1140 of Kluyveromyces lactis.
GN ORFNames=KLLA0E05225g;
OS Kluyveromyces lactis NRRL Y-1140.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=284590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neugeglise C., Tallia E., Goffard N., Franchin L., Aigle M., Anhouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykaeten C., Boissane A., Boyer J., Cattolico L., Confanioli F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H., Niclaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A., Sellenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts."
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382125; CAG99270.1;
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008565; F:protein transporter activity; IEA.
DR GO; GO:0006886; F:intracellular protein transport; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR006955; Uso1_p115_C.
DR InterPro; IPR006953; Uso1_p115_head.


```
Db 441 ELGENKIFEVEDLQGLIKLEVLDSNYIKDISLKNLTDIKELKLNKNKVSISIVANN 500
Qy 637 SHIAFNALDND-----GDEQFNKVVVKTTHNSYALADGE 672
Db 501 KNLQRLYINDNNVTTLYLKDAKDLVLTANNKITSFEGLELDDKEIH-----551
Qy 673 HFIVDPDKLS-----STIVDLEKI-----LKLIEGLDYSTLRQTQTOTPRD-----713
Db 552 ---VDNNKISKDPLKLNKLELTLSARTNVISDLKPIENLDYIKNLYLYENKISDISPLK 608
Qy 714 -MTTAGKALLSKNLR 728
Db 609 NMTGMLRLYLDKNKIK 624

RESULT 54
RBPI_PLAVB
ID RBPI_PLAVB STANDARD; PRT; 2873 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN Name=RBPI;
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92313338; PubMed=1617731; DOI=10.1016/0092-8674(92)90642-P;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merosomes";
RL Cell 69:1213-1226 (1992).
RN [2]
RP REVISIONS TO 1027-1036; 1110; 1278-1280; 1564; 1599-1607; 1804; 1861;
RP 1942 AND 2101-2103.
RA Tran T.M., Galinski M.R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
CC human reticulocyte cells.
CC -!- SUBUNIT: Homodimer (Potential).
CC -!- SUBCELLULAR LOCATION: Membrane-bound.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
Db 280 EYEINYAKRESKTETETLGDKETKLSAVSQAYAKHLESYKGVLPKPLKINDIKNNAFSVLKDS 339
Qy 129 PSTSNWEIC-DPITKGNLTVLGSLKSGVEKLQSOTDHLVLPSSQAADGTQLIQVASFATPPDK 187
Db 340 YCKDN---CGEVQKNTNKNFPISSSDQYKMEAVYIPKSIINDYTVLQKILS-----389
Qy 188 KTAIAEYTSRAGENGESIQDVGKEIINE-----GEVFNYSYLLKKVTIPTGYKHIG 239
Db 390 -----ESNELGIDIOETVNSLKLGBEISEVSHLYVINSTLIDDAAKKL- 433
Qy 240 QDAFVDNKNIAEVNIPESLETISDYAFAPHALAKQDLPNLKAIAGELAFDQITQKLSL 299
Db 434 -ESINEEDESAEDLQK-----FEDNSKALANNYCFIYI-----467
Qy 300 PROLMRLAERAPKSNHIKTIFRGNLSKLVIGEA--SFQNDLSQMLPGLGLEKIESEAF 357
Db 468 -KTLNEPIKAYESKVIKSNELLSTIIDTLGKSATALQESTFDQ-----ECCNKIKTEA- 520
Qy 358 GNPFGDDHYNRVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPE-----401
Db 521 -----EKKVDDAEDICEKN-----EQIYVEIPESEDETIDDKINDLQDL 559
Qy 402 IDYTKWLEEDFTYQKNSVTGFSNKGLOKVKRNKL-EIPKQNGVTITIG--DNAFRNV 458
Db 560 IDQKEYKDEIV---NNSEFISN-----RYKNIYENLKETYETELNDIGLENDTSKV 609
Qy 459 DFQNKTLRKYLDEEVKLPSTIRKIGAF-----AFQSNLKSFEA 497
Db 610 NFYLMQIRKINTEKTKIDESLQTVKFKYELDSKEIYELKIEFKSVTEINRLQDGS 669
Qy 498 SDDL--EIKEGAFMNNRIETLE-LKDKLVITGDAAFHINHIYAIVLPSVOEI-GRSAP 553
Db 670 ARDLHEEQIKE--ILDKMAKKVHLYKELLSLKGSSVYPTENNELLNTASYDNMGFSAK 727
Qy 554 RQGANNLIFMGSKVKTIGEMAFI-----SNLEHLDDLSEQKQLT-----IP 596
Db 728 KEKADNDNALNSVYREDINALIBEVEKFVTENKESTLEMLKDBEMEEKLODAKETPAK 787
Qy 597 VQAFSDNALKEVLLPASLKTIRREAFKN-----HLKQLEVASALSHIAFALDDN 647
Db 788 LNFVSDDKLTDTVTQMSAEVTVNAEGIKKIAQKQFENVHKKWKFSDAPS-TKFEALQNS 846
Qy 648 -----DGD-----EQFDNKVVVK-----THNSYALADGEHF 674
Db 847 MQQYNQEGDAIEKHKQNRSEKEEYFKNESVEEDLSRETEQEYTKHKQNSRRKGE---904
Qy 675 IVDPKLSTIVDLKILKILG-LDY-----STLR 704
Db 905 -----ISAEITNMRVINKIESQLNYGVIEKFSLIGDQNEVSTAKALKEKIVSDSLR 958
Qy 705 QTT---QTOPRDMTTA-----GKAL--LSKSN-----LRQGEKQK 734
Db 959 DKIDQYEFKEKTSVAVENTVSTIOSLSKAIDSLKRLNGSINNCKYNTDIDLLRSKIKT 1018
Qy 735 FLQEAQFFLGRVDDLKAKAKAEKALVTKK-ATPKNQQLLERSINKAVLAYNNSAIAKNAVYK 793
Db 1019 LREEVQKEIAETEGDKVVGENTTALLSLKLRDKGKINEK-----LNDGLNSLDTK 1070
Qy 794 RLEKELDLLTGLVEGKGPLAQATMVQGVYLLKTP-----PLPEYIIGLVNVPD-----842
Db 1071 ---KE-DLLKFYSESKSIHLKSKQKGP---QDPLNRIDEDWEDIKREVDLNNVQVISE 1123
Qy 843 -----KSKGLIYALDMS---DTIGEG---OKDAYGNPI-----LNVEDNEGYHALAV 884
Db 1124 NKVTLFKNSVTVIEAMHSHINTVAHITSNNKILSKVKEVEDKLNVLVEQNEDEYK--KV 1181
Qy 885 ATLADYEGID-----IKTILNSKLQSLTSIRQVPTAAVHRA-----920
Db 1182 KNPENEKQLEAIRGSMKLEKVINKHVEMQLESTANTLKSNAKENEHLEBLNKTK 1241
Qy 921 ----GIFOAIQNAAAEAEQLPKPQTHSEKSSSSSSSANSKORGLQSN 963
Db 1242 GQMRDIYEKLIKIAEEL-----KEGTVELKDANEKANKVELEFERN 1283
```

```

RESULT 55
Q8RGS5 PRELIMINARY; PRT; 197 AA.
ID Q8RGS5;
AC Q8RGS5;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical cytosolic protein.
GN OrderedLocusNames=FN0209;
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RX DOI=10.1128/JB.184.7.2005-2018.2002;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Busch G., Haselkorn R.,
RA Ponstein M., Kyrpides N.C., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AB010533; AAL94415.1; -.
KW Complete proteome.
SQ SEQUENCE 197 AA; 22356 MW; 460289F03260F5A4 CRC64;

Query Match 4.1%; Score 208; DB 2; Length 197;
Best Local Similarity 25.6%; Pred. No. 0.062;
Matches 61; Conservative 43; Mismatches 74; Indels 60; Gaps 4;

QY 402 IDYTWLEEDFYQKNSVTGFSNKGKQVK--RNKNLSIPK-QHNGVTITEIGDNAPRV 458
DB 1 MDONTWEYDDDFKGDDELKGMTQKQKQKVEGKTDLVPELTPDGLPLKXIADNAF--- 57

QY 459 DPMKTLRKYLEEVKLSPIRKIGAFAPQSNLKSFEASDDLEIKGAFMNNRIETLE 518
DB 58 -----YRGLTSVIISTVESIGYDAFVCKLKEVKLPEALVNIIEGFAFRNKLTKVE 110

QY 519 LKDLVTIGDAAPHNHIAIYVLPSVQIEIGSAFRQGANNLIPMGSKVKTGLGEMAFLS 578
DB 111 FGSKVKRLPSPSFMNLSLNLPTVEYIGASAFVK----- 147

QY 579 NPLEHLDLSEQQLTEIPQAFSDNALKEVLLPASLKTIREAPKKNHLKQLEVASAL 636
DB 148 -----NSLETVPSPKSVTKIDMYAFRNKNIHKVEVANSV 181

RESULT 56
Q64VN1 PRELIMINARY; PRT; 520 AA.
ID Q64VN1;
AC Q64VN1;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Putative cell surface antigen.
GN ORFNames=BF1698;
OS Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YCH46;
RA Kuwahara T., Yamashita A., Hirakawa H., Nakayama H., Toh H., Okada N.,
RA Kuwahara S., Hattori M., Hayashi T., Ohnishi Y.;
RT "Genomic analysis of Bacteroides fragilis reveals extensive DNA
RT inversions regulating cell surface adaptation."

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2004).
DR EMBL; AP006841; BAD48445.1; -.
SQ SEQUENCE 520 AA; 57220 MW; 17CF3336B968925F CRC64;

Query Match 4.1%; Score 208; DB 2; Length 520;
Best Local Similarity 21.5%; Pred. No. 0.24; Mismatches 91; Indels 148; Gaps 22;
Matches 121; Conservative 91; Mismatches 203; Indels 148; Gaps 22;

QY 49 ESSGSKLKINETSGPVDVTDLFSKRTTPPEKIKONLAKGPREQELKAVTENSEKQ 108
DB 42 ELPGRYFIYFGEELDVSDMTLKVFSNSETSIVSKXDEITGPN-----NTVPEN-DQILE 96

QY 109 ITSQSQLEBQSKESLSLN-----KTVPSSTNWEICDPITKGNLVLGSLKSGVKEKLSQTD 161
DB 97 VHKSFTVTFKIQLINDIQAISIKTLPSKTVTYTLGPELSLN----- 139

QY 162 HLVLPSQAADGTQLIQVASFAPTPDKKTAIAEYTSRAGEGEISQLDVDGKEIINEGRV- 220
DB 140 -MWLEINVADGTIKENSAPSA-----DWQGFNSVPAQQLVTLDDGKQVDFVOIL 192

QY 221 -----FNSYLLKKVTIPTGYKHHGQDAFVDNKNI--AEVNLPSLETIS 262
DB 193 PVKVDGDKVWSVIDSDFTS-----ITPDDGIRTIGSKAF-ENKNIKASELLFPASLSTIE 246

QY 263 DYAFAHIA-LKQIDLP-DNLKAIGELAFFNQITGKLSLPRQLMLRAERAPKS----- 313
DB 247 QAAPAYCRNLKIVDLSHTSIKELPEEAPLFSGIK-KIALPASLRIVGKEAFYGCCTDLNVI 305

QY 314 --NHKITEPRG-----NSLKVIGASP-QQNDLSQMLPDGLEGKIESAP 356
DB 306 DISHTSVKELQNGAFGKSGISSISLPSTFKLVGTSAFIETKNLKEILLTPESEVIDLEAP 365

QY 357 TGNPGDDHYNRRVVLTKSGKNPSGLATENTYVNPDKSLMQWESPEIDYTWLEEDFYQK 416
DB 366 -----SGSSIQKVTLPNTIYHIDRSFY-NCPELITIE-----TYGT 400

QY 417 NSVTGFSNKGKQKVRKNKLEIPKHNGVTITEIGDNAPRVDPQNTLRKYDLEVKLP 476
DB 401 RTTFS-----PVDRTAAIVSECFNHPK-----LTVLKIP 430

QY 477 STIRKIGAFQSNLKSFEASDDLEIKGAFMNNRIETLEKDKLVITIGDAAPHNH 535
DB 431 ASIAKIGISALNCKQKVTLLIPVSVKALDFNAFNAVSLDEISLMSPTMTADYYPVPPR 490

QY 536 IYAIVLPSVQIEIGSAFRQNGA 558
DB 491 IQKIRVPQNLVE---TVKQNK 509

RESULT 57
Q839L8 PRELIMINARY; PRT; 901 AA.
ID Q839L8;
AC Q839L8;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Surface exclusion protein, putative.
GN OrderedLocusNames=EF0146;
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927; DOI=10.1126/science.1080613;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seehadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L.A., Brinkac L.M., Beanan M.J.,
RA Dougherty S.C., DeBoy R.T., Durkin S.A., Kolonay J.F., Madupu R.,
RA Nelson W.C., Vamathevan J.J., Tran B., Upton J., Hansen T., Shetty J.,
RA Khouri H.M., Utterback T.R., Radune D., Ketchum K.A., Dougherty B.A.,
RA Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant

```



```

QY 840 --YFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYH 880
    |||::|||::
Db 1831 QKISDKKNKFLHALNET-----PIPNFNTLKEIYH 1860

RESULT 60
Q750B4
ID Q750B4 PRELIMINARY; PRT; 3645 AA.
AC Q750B4;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE AGR043WP.
GN ORFNames=AGR043W;
OS Aeshya gossypii (Yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10895;
RA Lerch A., Brachat S., Voegeli S.E., Gaffney T., Philippsen P.,
RA Dietrich F.S.; (2004) to the EMBL/GenBank/DBJ databases.
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016820; AA554532.1; -.
DR AGD; AGR043W; -.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00233; PH; 1.
SQ SEQUENCE 3645 AA; 409727 MW; 839C556DC514E73A CRC64;

Query Match 4.0%; Score 205.5; DB 2; Length 3645;
Best Local Similarity 20.0%; Pred. No. 4.6;
Matches 239; Conservative 175; Mismatches 434; Indels 347; Gaps 54;

QY 5 LKTVALTLTVSVVTHNQEVFLVKEPIILKQTAQSSISGADYAESSGSKLKINETSGP 64
    |||::|||::
Db 494 LQSGGVKVTISEYEALSEKANLVSAPPLEFLE-----EQAKNIGKVIIDQEEYASL 545

QY 65 VDDTVTDLSDKRTTPEKTKDNLAQPREQELKAVNTESKQITSGSOLQO-SKESUS 123
    |||::|||::
Db 546 QDPVSVDTIISKAESHG---FDVLSKSAHQMLKSIQPTLEYLQ----SKLQHQSMKAIK 598

QY 124 LKTVPTSTNWEICDPTIKGTNTLVGLSKSGVEKLSQTDHLVLPSSQAADGTOLIOVASPAF 183
    |||::|||::
Db 599 LNDYALFENPTE--DYLSSKARL--LEKS-----LVNSSELAELURKISQBPSPF-- 642

QY 184 TPDKTAIAEYTSRAGENGIEISQLDVKGKEIINEGEVFNLSYLLKKVTIPTGYKHI----- 238
    |||::|||::
Db 643 -----DYLSSRAAANTSCQLLPDDQFDALN-----KKANSPT-LEHLKSASA 682

QY 239 GQAFVDNKNIAEVLNPESETISDYA--FAHLAKQIDLPNLKAIGELAFDQITGK 296
    |||::|||::
Db 683 ALDHVLVDKDEYHLNPPLEKLADLAATYKHALSQAHEHSLHPSAEL-----LTKR 736

QY 297 LSLPRLMLAER-----AFKSNHKT-----EPRGNSLKVI----- 329
    |||::|||::
Db 737 LTELQVMLPKKEYENLLSPTKHDIQYAAACHILVSDDEPKQLTSPSIVVIOQASL 796

QY 330 GEASFQNDLSQLMPDGLKEIESE-----APT-----GNPGDDHYNRVVL 371
    |||::|||::
Db 797 AHAVPNEEELAKHPD-LETIKSHLSEIATYALPIEELGNMKQISNPTLHIS----- 850

QY 372 WTAKGNKPSGLTENTYVNDPSLWQES-PEIDYTK-----WLSE--DFTVQ 415
    |||::|||::
Db 851 -AKAQHQSVLLPIDTY---NKLLEDSDPSMDVMVKKHTEKDKKELLASLAQQLDSHIM 906

QY 416 KNSVTGFSKNGI-----QKVENKLEIPKQHNGVTIITEIGDAFRNVDQNKTLR 466
    |||::|||::
Db 907 KNSHAELLKRLDSPTLEFLAKAKAYDILLPEKKEHDIILLSTMENPA---IDYLQTKAK 963

QY 467 KYDLBEVLKLPSTIRKIGAFAPQSNLKLSPASDLEIEKEGAFMNNRIETLE----- 518
    |||::|||::

```

```
DR Pfam: PF01468; GA; 34.
KW Complete proteome.
SQ SEQUENCE 9439 AA; 1050776 MW; 5A8C6B8E9AA31CDD4 CRC64;

Query Match
  4.0%; Score 205; DB 2; Length 9439;
Best Local Similarity
  19.1%; Pred. No. 18;
Matches 201; Conservative 178; Mismatches 447; Indels 224; Gaps 45;

QY 47 YABSSGSKLKINETSQGV--DDVTDLFSKRTTPEKIKONLAKGPREQELKAVTENTE 104
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6854 YQAVNKAKNIINDQPTPMANDEIQSVLNEVKQT---KDNL-----HGQKLANDKDT 6904

QY 105 SEKQITSGSQLEOSKESLSLNTKVTSTNWEICDFITKGNLTGLSKSGVEKLSQTDHLV 164
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6905 AQATLNALYNLQAQGNLETKVQNSRPEVQKVQLANQLNDAMKDLDAITGNDIAIK 6964

QY 165 LPSQAADGTQLIQVAFAPDKKTAIABYTSRAGENGESISOL-----DVGDEIIL 215
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6965 QTSYINEDTSQVNFDEYTDGKNIIVAEQTPNPNMSPNTNIADKIITEAKNDLHGQKL 7024

QY 216 NEGEVFNYSLLKKVT-IPGYKHIGQDAFVNKNIAVFN-----LPSELETISDYAF 266
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7025 KQAQOQSINTINQMTGLNQAQKEQLNQBIQOTQTRSEVHQVINKAQALNDSMTLRQSIT 7084

QY 267 AHLALKQIDLPNLKAIGELAFPDN-----QITGKLSLPQMLAERAFKSNHIKTIE 320
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7085 DEHEVKQTSNYIN-ETVGNQTAYNNAVDRVQIINQTSNPTNPLEVERA-----TSNVK 7138

QY 321 FRGNSLKVIGASFOQNDLSQMLP--DGLKIESEAPFG-----NPGDDHNNRV 369
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7139 ISKDALH--GERELNDNKNSTPAVNHLDNLNQAQKEALTHEIQATIVSQVNNIYNKAK 7196

QY 370 VLWTKSGK-----NPSGLATENTYVNP-----KSLMQE-----SPEIDYTKW 407
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7197 ALNNDMKKLLKQIVAQDQVNSNNYINEDSTPNNYNDTINHAQSIIDQVANPMTSHDE- 7255

QY 408 LEEDFTYOKNSVTGFSNK-GLQVKRNKNLEIPKQHGNTTITEIGDANFRVDFQNKTLR 466
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7256 IENAINNTKHAINLGDGHKKQQAQENANL-----LNSLDNLNAPORDAINRLVN 7306

QY 467 KYDLEEVKLPSTIRKIGAFAPQSNLKSFEASDDLEIEKEGAFNMNRITETLEKDKLVTI 526
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7307 EAQTFE-----KV-----AEQLQSAQALNDAMKHLRSIQNQ--SSVRSQSKYINA 7350

QY 527 GDA-AFHNIHIAI-----VLPEVQEIIGRSAPRO-----NGANNLIPMGSKVTLGEMA 575
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7351 SDAKKEQYNH--AVREVENIINEQHTLDEIKETIKQITDGVNQANNDL-----7395

QY 576 FLSFLEHLDLSEQKQLTEIP-----VQAFSDNALKE-----VLLPAS 613
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7396 ---NGVELLDADKQNAHQSIPTLMHLNQAQNALNEKINNNAVTRTEVAIIIGQAKLLDHA 7452

QY 614 LKTRREAFKKNHLKQLEVASALSIAFNAL--DDNDGDGEQPDN-----KVVVKTHTNSY 666
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7453 MENLEESIKDEQVQK-----SSNYINEDSDVQETVDYDNAVHYTEILNQTVNPTL 7502

QY 667 ALADGEHFIYVDPDKLSSTIVLEKILKLIIEGLD--YSTLRQTFQTFQDMTTAGKALLSK 724
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7503 STIEDIEHAINEVQAQKQLRGKQLQYITIDLADKELSKLDLTSQQSSISINQIYAKTR 7562

QY 725 SNLRQG--EKQFLOBAQPLGRVDLDKAIKAERKALVTKATKNGQLLERSINKAVLAYN 783
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7563 TEVAQAIERAKSLNHAM-----KALNKVYKNADKVLDSRFINEDQPEKKAYQQAINHVD 7617

QY 784 NSAIKKAN-----VKRLEKELDLTLGLVBEKGKPLAQ-----ATMVQGVYLLKTPPLP 831
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7618 SIIHQTPNDPTVINSTHETETAQNHLGDQKLAHAQQDAANVING--LIHLNVAQR 7675

QY 832 EYIIGLNIVYFKSGKLIIVALDMSDTIGG-----QKDAYGNPILN-----VDEDEGYHAL 882
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7676 EYMIINTNTATTREKVAKNLNAQAALDKAMETLQQVVAHAKNNILNDSKYLNEDESK-YQQQ 7734

QY 883 AVATLADYEGLDIKTILNLSQLSITSIRQVPTAAVHRAGIFQAIQNAABAEQQLPKPGT 942

Db 7735 YDRVIADAEQL-----LNQTTN-----PTLEPKVDIVK--DNVLANEKILF----- 7774
QY 943 HSEKSSSESANSDKRGLOSNPKTNRGRHS 972
Db 7775 GAELKSYDKS-NANDEIKHMYLNNAAQKQS 7803

RESULT 62
Q73CV3 PRELIMINARY; PRT; 1508 AA.
AC Q73CV3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE LPXTG-motif cell wall anchor domain protein.
GN OrderedLocusNames=BCCE0961;
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714; DOI=10.1093/nar/gkh258;
RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA Nelson W.C., Kistner A., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RL Nucleic Acids Res. 32:977-988(2004).
DR EMBL; AE017267; AAS39892.1; -.
DR TIGR; BCCE0961; -.
DR InterPro; IPR008454; Cna_B.
DR Pfam; PF05738; Cna_B; 11.
KW Complete proteome.
SQ SEQUENCE 1508 AA; 164822 MW; EC32A5763C184649 CRC64;

Query Match
  4.0%; Score 204.5; DB 2; Length 1508;
Best Local Similarity
  19.9%; Pred. No. 1.5;
Matches 246; Conservative 144; Mismatches 444; Indels 401; Gaps 56;

QY 13 TTTSVVTNHOEVFLSKPEILKQTOASSISGADYAESSGK-----SKLKINETSGPVDDT 68
DB 413 TPVKVTIDNTNIIIEFVKNTKVENQPKLLKXD--SESQGLLPKAFDVIDKGVVET 470

QY 69 -VTDLFSKRTTPEKIKONLAKGPREQELKAVTENTESEKQITSGSOLESKES--LSLN 125
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 471 IVTD-----DKGALSKQLPVGSYTLKEVEAPKGYELSSSVSDVVEVN 514

QY 126 KTVPTSTNWEICDFITKGNLTGLSKSGVEKLSQTDHLVLPQQAADGTQLIQVAFAPTP 185
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 515 KVV-----TVDVVNNKIPKVTQGFIV-KVDAEDTKVLSDAEFVEYK 557

QY 186 DKTAIAEYTSRAGENGESISOLDVDGKEIINEGVEFNSYLLKVKVTIPTGYKHIGQDAFVD 245
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 558 DGKQVDTLRTDKTK--VVSQKLEPGK-----YTLKETKAPQGYKLLKEEIEV- 603

QY 246 NKNTAEVNLPESETISDYAPAHALAKQIDLPDLNKAIGELAFPDNQITGKLSPLRQLMR 305
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 604 ---VVEAN-----KVVQVQV-ENAKELGSL-----QVTKKDAESSGVLE 638

QY 306 LAERAFKSNHIKTIEPRGNSLKVIGASFOQND-----LSQLMLPDGLEKI 351
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 639 GAE-----FRLKNENGQVVGGETTKTNKQGVKFNVLVPOKYTLEETKAPGEGYKAV 688

QY 352 E-----SEAFNGN-----PGDDHNNRVVLT-----KSGKNPSPGLAT 384
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 689 EVTVVNVVANEVVKQEVNNEKLTGQFEIVKVDKDAEDKAKVLSDAEFVEYKQKVKETLRT 748

QY 385 ENT-----YVNPDKSLWQSPSEIDYTKWLEEDF--TYQKNSVTGFSNK-----GLQKV 430
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 749 DKTGKVISQKLEPGTYTLKETKAPQGYKLLKEEIEVVVEANKVVEVQVENAKELGSLQVI 808

QY 431 KBNQNLIPKQHNGV-----TITEIGDANFRVND-----FQNKTLRYDLEEVKLPSTI 479
```

```
Db 809 K--KDAESGKVLGAEAFKLNKNETGQVVGTTKTKNGDGVVVKFENLVPKGYTLBETKAPEGY 866
QY 480 RKIGAFAPFNNLKSPEASDDLEEIKEGAFMNRIB-----TLELKDKLVITIGDAAFHI 533
Db 867 KAVEV-----TVEVNVVAVVQVEMKEKVTGQFEIIVKVDANDKAKVLSDAEFTV 917
QY 534 NHIYAIVLPSVQIEGRSFRQNGANNLIIFMGSKVKTGEMAF-----LSNRLSH 583
Db 918 YKDGKKVAELKTDSECK-----VMSPKL-PLGEYTVKTKAPEGYKLSNKEWK 964
QY 584 LDLSEKQLTETPVQAFSDNALKEVLLPASLKTIREEAFKQKHLKOLEVASLSHIAFNA 643
Db 965 VTIQNEKEV--VKVEAENERIL-----GSLQIITKD--DKQAKLSLGAEP-----T 1007
QY 644 LDNDQDE-----QPDNKKVVVTHNSYALADGEHPIVD-----PDKLSST-----IVDL 689
Db 1008 LKDAQNVVKEGITTDKSIGKVD-----GLVGEYTLBETKAPEGYELTKQVHVHTVDGE 1063
QY 690 KILKLEGLDYSTLRTOTTQTFQDMTTAGKALLSKSNLRQGEKQKPLQBAQFFLG-----744
Db 1064 KI-----VDVKVTSKSLGQFEIVKVD-----DKAKVLSDAEPEVYKDGKK 1106
QY 745 -----RVD-LDKAIA-KAEKALVTKATKNGQ-----LLERSINKAVLAYNNSAIKKANVKRL 795
Db 1107 VETLRDTKTKGVISOKLEPGTYTLKETKAPQGYKLLKEBIEVVVVEANKVQVQVENAKEL 1166
QY 796 -----EKELDLTLGLVEGK-----GPLAQATWYQGVYLLKTPLP 829
Db 1167 GSLQVKKDAESGKVLGAEAFKLNKNETGQVVGTTKTKNGDGVVVKFENLVPKGYTLBETKAPEGY 1225
QY 830 LPSEYIIGLVNYPD-----KSGKLIIVALDMSDT-----IGRG 860
Db 1226 APEGYKALEVTVVNVVANTVIKQVLEKVEKIKGVQVEITKVATDNTKKLAGAVPEI 1285
QY 861 QKDAYGNPILNVDENEGHALAVATLADY-----EGLDIK-----TILNSKLS-- 904
Db 1286 LKD--GTKIDTLTDDKNGKATSKLEPGDYILKEVQAPGEGYELSDKGIEFTISNOKIEV 1343
QY 905 --QLTSIRQVPTAAVHRAGIFQA-----IQNAABAE-----QLLPKGEYHSEK-- 946
Db 1344 KLOITNEKTSKGPENPGGETETPGGETETPGGETETPGGETETPGGETETPGGETETPGGETETPG 1403
QY 947 -----SSSESANSKDRGLQ 961
Db 1404 BETKPGEBETKPGEBETKPGEBETKPGEBETKPGEBETKPGEBETKPGEBETKPGEBETKPGEBET 1463
QY 962 SNPKTNRGHRSAILPRTGSKGSFVYGLGYTSVAL 996
Db 1464 TLPERKQOGASHAQLPATGHDMNVL-PFIGFALVLL 1497

RESULT 63
Q99QR6 PRELIMINARY; PRT; 2481 AA.
AC Q99QR6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE FmtB protein.
GN Name=fmtB#mrp; OrderedLocusNames=SAV2160;
OS Staphylococcus aureus (strain Mu50 / ATCC 700699).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699;
RX MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
RA Kuroda M., Ohba T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
```

```
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus";
RL Lancet 357:1225-1240(2001).
DR PIR; D90011; BAB58322.1; -.
DR EMBL; AP003364; BAB58322.1; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR000890; Acetate_kin.
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR005877; Gpos_Ysirk.
DR InterPro; IPR0011490; FIVAR.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF07564; DUF1542; 18.
DR Pfam; PF07554; FIVAR; 1.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 2481 AA; 263765 MW; ELEAB99B81665E8 CRC64;

Query Match 4.0%; Score 204.5; DB 2; Length 2481;
Best Local Similarity 18.5%; Pred. No. 3;
Matches 206; Conservative 159; Mismatches 403; Indels 343; Gaps 45;

QY 10 LTLTIVSVTHNOEVPSLVKBPILKQ-----TOASSSI-----SGADYAESGSKLKI 58
Db 1293 LQATQTNVINNDQATNEEKEAAIQQLATAVTDKANNITAAATDDGVDGTADKAGNSIQS 1352
QY 59 NETSGFVDDTVTLPSDKRTTPEKIDKLAKGPREQELKAVTENTSEKQITSGSOLEQS 118
Db 1353 TOPATAVKSNAKNEVDQAVTTONQAIQNTT-----GATTEKNAKDLVLKAKEKAY 1404
QY 119 KESISLANKT--VPSTSNWEICDF--ITKGNLTVLGSLKSGVEKLSQTDHLVLPSSQAADGTQ 174
Db 1405 QDILNAQTNDVDTQIKDQAVADIQGITADTTIKDVAK-----DELATKAN-----1449
QY 175 LIQVASFAPTPDKKTAIAEYTSRAGENCEISQLDVGKKEINEGEVFNFSYLLKKVTIPTG 234
Db 1450 -----EQKALIAQTADATTEKEQANQVDAQ-----LTQG 1480
QY 235 YKHIGQDAFVNKNIENLPSLETISDYAFALHALAQIDLPDLNKAIGELAPFDNOIT 294
Db 1481 NQNTENASQIDVNTAKDNAIQADPIQ-----ASTDVKTNARA-ELLTENQNKIT 1530
QY 295 GKLSLPLQMLRAERAPKSNHIKTIEPRGSLKVTIGEASFQNDLSQLMLPDGLEKIBSE 354
Db 1531 EIL-----NNNETTNEEGNDIGPV-RAAYE-----EGLNNINRA 1564
QY 355 AFTGNPGDDHNNRVNVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLBEDFTY 414
Db 1565 TTTGD-----VTTAKDTAVQVQQLHANPVKPKGAK-----KELDQAAAD 1604
QY 415 QKNSVTGFSKGLQKVENKKNLEIPKQHNGVTIIEGDNAFNVDFQNKTLRKYLEEVK 474
Db 1605 KKTQIEQTPNASQQEINDAKQ-EVDTELN-----QAKTNVD-QSSTNEYVD-----1648
QY 475 LPSTIRKIGAFAPFQSNLKSFEASDLEIEIKEGAFMNNRIETLE--LKDKLVTIGDAAFH 532
Db 1649 -----NAVKEGRKAKINAVKTFSEYKKDALAKIEDA---1678
QY 533 INHIYAIVLPSVQIEGRSFRQNGANNLIIFMGSKVKTGEMAFISNRLSH-----DL 586
Db 1679 -----YNAKVNEADN-----SNASTSSEIAEAKOKLAELKQTDADQNV 1715
QY 587 SEQKQLTETPVQAFSD-NALKEVLLPASLK-----TIREAEFKK 624
Db 1716 NQATSKDDIEVQIHNDLNDINIDYITPTGKGSATTDLYAYADQKQKNTSADNTATQDBKQ 1775
QY 625 NHLKQLE--VASALSHTIAFNALDNDGDG-----QFONKVVVYKTHNSYALA- 669
```

Db 1776 QAIKQVQDNVQTALENIN-NGVDNGDVEDDALTQGKAADITIQVDATVVKPQAKAIEAKAE 1834
 Qy 670 DGEHFIIVDPDKLSS--TIVDLKILKLEGLDYSLRQTQTQFRDMTTAGKALLSKSNL 727
 Db 1835 DTKEIDHSDQLTAEKTEALAMIKQITDQAKQGITDATTAEVEKAKAQGLEAFDNIQI 1894
 Qy 728 RQGEKQKFLQBAQFLGR---VLDLKAIAKAEKALVT-----KKATKNGQLLERSIN 776
 Db 1895 DSTEKQKAELELETALDQIEAGVNVVDADATTEKEAFNALEDILSKATED--ISDQTN 1952
 Qy 777 KAVLAYNNSAIKANVVKRL-----EKELDLTLGLVEGKPLAQATMVQ 820
 Db 1953 AEIATVKNLSALEQLKQINPVVKNKALEIAEVVVKQIEII-----KNADADASAKE- 2005
 Qy 821 VYLLKTPLEPYIYGLNVYFDK-SGKL-----IYAL-----DMS 854
 Db 2006 --IART-----DLGRYDFRPAKLDKTQNTTEVAELQNVITPAIEAIVPQNDPDAN 2054
 Qy 855 DT-IGEGOKDAYGNPILNVDEDEGHALAVATLADYEGLDIKTILNKLSQLTSIROVP 913
 Db 2055 DTNNGTDNDATNSANATPENTQPNVSETT--DNKADASPPTPNNSDAATGETTVT 2112
 Qy 914 TAAVHRAGIFQAIQIAAAAAEQLLP-----KPGTHSEKSSSES 952
 Db 2113 SATDDAKDKPQANNSSADASTNSPTMDNDVTSKPEVESTNGTDDKPVTTETDNATPAES 2172
 Qy 953 A---NSKORGLOSQNPKNRHSAILPRTGS 980
 Db 2173 TTNNSSTTTATNENAPTQ---STATAPTAS 2200

RESULT 64

Q7A4B1 ID Q7A4B1 PRELIMINARY; PRT; 2481 AA.
 AC Q7A4B1;
 DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
 DE FmbB protein;
 GN Name=fmbB(mrp); OrderedLocusNames=SA1964;
 OS *Staphylococcus aureus* (strain N315).
 OC Bacteria; Firmicutes; Bacillales; *Staphylococcus*.
 OX NCBI_TaxID=158879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
 RT "Whole genome sequencing of methicillin-resistant *Staphylococcus*
 RT *aureus*."
 RL Lancet 357:1225-1240(2001).
 DR EMBL; AP003136; BAB43253.1; --
 DR GO; GO:0009986; C:cell surface; IEA.
 DR GO; GO:0016020; C:membrane; IEA.
 DR InterPro; IPR000890; Acetate kin.
 DR InterPro; IPR011439; DUF1542.
 DR InterPro; IPR011490; FIVAR.
 DR InterPro; IPR005877; Gpos_YSRK.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR Pfam; PF07564; DUF1542; 18.
 DR Pfam; PF07554; FIVAR; 1.
 DR Pfam; PF04650; YSRK signal; 1.
 DR TIGRFams; TIGR01167; LPTXG anchor; 1.
 DR TIGRFams; TIGR01168; YSRK signal; 1.
 DR PROSITE; PS01075; ACETATE_KINASE_1; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 2481 AA; 263765 MW; E1EAB99B81665E8 CRC64;

Query Match 4.0%; Score 204.5; DB 2; Length 2481;
 Best Local Similarity 18.5%; Pred. No. 3;
 Matches 206; Conservative 159; Mismatches 403; Indels 343; Gaps 45;
 Qy 10 LTLTVSVVTHNQVFSVLVEPILKQ-----TOASSSI-----SGADYAESSQSKLKI 58
 Db 1293 LQATQTVINNDQATNEEKEAAIQQLATATVDKNNIATATDDNGVDVDTAKDCKNSIQS 1352
 Qy 59 NETSGPVDVTDLPSKRTTPPEKIKONLAKGPQEBELKAVTENTESEKQITSSQLEQ 118
 Db 1353 TQPATAVKSNKNEVDQAVTTQNOAIDNTT-----GATTEKNAAKDLVLAKEKAY 1404
 Qy 119 KESISLAKT--VPSTSNWEICDP--ITKGNLTGLVSGVGEKLSQTHLVLPSQAADGTQ 174
 Db 1405 QDIUNAQTTNDVTQIKQAVADIQGIATDTTIKDVAK-----DELATKAN----- 1449
 Qy 175 LIQVASFAPTPDKKTAIAEYTSRAGEISQLDVGKEIINEGEVFNYSLLKKVTPTG 234
 Db 1450 -----EQALLIAQTADATTEKEQANQVDAQ-----LTQG 1480
 Qy 235 YKHIGQAFVQDNKNAEVLNLESLETISDYAFALHALKQIDLPNLKAIGELAFDNIQIT 294
 Db 1481 NQNIENAQSIDDVNTAKDNAIQADIPQ-----ASTDVKTNABA-ELLTEMONKIT 1530
 Qy 295 GKLSLPQLMELAEAFKSNHIKTIETFRGNSLKVIGRASFOVDNLSQLMLPDGLEKTESE 354
 Db 1531 EIL-----NNETTNEKGNIDIGPV-RAAYE-----EGLNNINAA 1564
 Qy 355 AFTNGPGDDHYNNRVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKMLEEDFTY 414
 Db 1565 TTTGDD-----VTTAKDTAVQKVQQLHANPVKPKAGK-----KELDQAAAD 1604
 Qy 415 QKNSVTGFSKGLQKVKRNKLEIPKOHNGVTITEIGDNAPRVNDFQNTLURKYDLEEVK 474
 Db 1605 KKTQIEQTPNASQOEINDAKQ-EVDTELN-----QAKTNYD-QSSTNEYVD----- 1648
 Qy 475 LPSTIRKIGAPAFQSNLKSPEASDLEEKIEKGFPMNRIETLE--LKDKLVITIGDAAFH 532
 Db 1649 -----NAVKEGKAKINAVKTFSEYKDALAKIEDA--- 1678
 Qy 533 INHIYAVLPESVQIEGRSAPFRONGANNLIPMGSKVTTLGEMAFLSNRLEHL-----DL 586
 Db 1679 -----YNAKVNEADN-----SNASTSSSIAEAKQKLELKQTADQNV 1715
 Qy 587 SEQQLTEIPVQAFSD-NALKEVLLPASLK-----TIREAPKK 624
 Db 1716 NQATSQKDDIEVQIHNDLDNINDYTIPTGKKESTTDLAYADQKKNISADTNATQDEKQ 1775
 Qy 625 NHLKOLE--VASALSHAFNALDNDGDGE-----QFDNKVVVTKHNSYALA- 669
 Db 1776 QAIKQVQDNVQTALENIN-NGVDNGDVEDDALTQGKAADITIQVDATVVKPQAKAIEAKAE 1834
 Qy 670 DGEHFIIVDPDKLSS--TIVDLKILKLEGLDYSLRQTQTQFRDMTTAGKALLSKSNL 727
 Db 1835 DTKEIDHSDQLTAEKTEALAMIKQITDQAKQGITDATTAEVEKAKAQGLEAFDNIQI 1894
 Qy 728 RQGEKQKFLQBAQFLGR---VLDLKAIAKAEKALVT-----KKATKNGQLLERSIN 776
 Db 1895 DSTEKQKAELELETALDQIEAGVNVVDADATTEKEAFNALEDILSKATED--ISDQTN 1952
 Qy 777 KAVLAYNNSAIKANVVKRL-----EKELDLTLGLVEGKPLAQATMVQ 820
 Db 1953 AEIATVKNLSALEQLKQINPVVKNKALEIAEVVVKQIEII-----KNADADASAKE- 2005
 Qy 821 VYLLKTPLEPYIYGLNVYFDK-SGKL-----IYAL-----DMS 854
 Db 2006 --IART-----DLGRYDFRPAKLDKTQNTTEVAELQNVITPAIEAIVPQNDPDAN 2054
 Qy 855 DT-IGEGOKDAYGNPILNVDEDEGHALAVATLADYEGLDIKTILNKLSQLTSIROVP 913
 Db 2055 DTNNGTDNDATNSANATPENTQPNVSETT--DNKADASPPTPNNSDAATGETTVT 2112
 Qy 914 TAAVHRAGIFQAIQIAAAAAEQLLP-----KPGTHSEKSSSES 952

```

Db 2113 SATDDAKDPQANNSSADASTNPTMDNDVTSKPEVESTNNGTDDKPVTTEDNATPAES 2172
QY 953 A---NSKDRGLQSNPKTNRGHSAILPRTGS 980
Db 2173 TTNNSTTTATNENAPTQ---STATAPTTAS 2200

RESULT 65
Q9ZHL0
ID Q9ZHL0 PRELIMINARY; PRT; 4919 AA.
AC Q9ZHL0; Q7BY44;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE Large supernatant protein 2.
GN Name=lapA2; OrderedLocusNames=HD1156;
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000;
RX MEDLINE=99030326; PubMed=9811662;
RA Ward C.K., Lumley S.R., Latimer J.L., Cope L.D., Hansen E.J.;
RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like
RL protein."
RL J. Bacteriol. 180:6013-6022 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=35000HP / ATCC 700724;
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT "The complete genome sequence of Haemophilus ducreyi."
RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF057696; AAC9761.1; -.
DR EMBL; AE017154; AAP96012.1; -.
DR PIR; T31105; T31105.
DR GO; GO:0004197; P:cyteine-type endopeptidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008619; Fil_haemagg.
DR InterPro; IPR008638; Haemagg_act.N.
DR InterPro; IPR003951; Peptidase C58.
DR InterPro; IPR006473; Peptidase_C58_yt.
DR Pfam; PF05594; Fil_haemagg; 10.
DR Pfam; PF05860; Haemagg_act; 1.
DR Pfam; PF03543; Peptidase C58; 1.
DR TIGRFAMs; TIGR01901; adhes_NPXG; 1.
DR TIGRFAMs; TIGR01586; yopT_cys_prot; 1.
KW Complete proteome.
SQ SEQUENCE 4919 AA; 542602 MW; 5779201455CA69A0 CRC64;

Query Match 4.0%; Score 204.5; DB 2; Length 4919;
Best Local Similarity 20.0%; Pred. No. 7.8;
Matches 250; Conservative 181; Mismatches 415; Indels 405; Gaps 65;

QY 10 LTLTVSVVTHNQEVPSLVKEPIILKQTQA-----SSISGADVAESGSKLK----- 57
Db 598 VTLSKSTLSAGELTFPKYKNTLNDSELAANNLSLNASHNVTLNKKSLSAQAQADIK 657
QY 58 -----INETSQVDDTVTDLFSKRTTPEKI-----KDLN--- 87
Db 658 VNLTNDTTELTAKNLDINSTITNNGIAGIFAN--ITTEKLNKKEKALILAEQNLNFT 715
QY 98 -----AKPREQELKAVNTESKEQITSGSQLEQSKESLSLN----- 125
Db 716 VNGSHVENKGDIVSKDQATVTFESKNSDFTSNGSKLVNAQNQLKVNANNFTISQGDITLI 775
QY 126 -----KTVPTSNWEICDFTTKNTLVG-----LSKSGVB---KLSQT 160
Db 776 GNVTLNASGTFNTSGNLTTVKTLDVGDIQNFTNKGNTLVGDLHIKSKTKITNDGKLISI 835

```

RESULT 66
Q7RTL4
ID Q7RTL4
AC Q7RTL4;

PRELIMINARY; PRT; 17903 AA.

DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
TITIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20517449; PubMed=11062264; DOI=10.1083/jcb.151.3.639;
RA Machado C., Andrew D.J.;
RT "D-Ritin: a giant protein with dual roles in chromosomes and
RT muscles";
RL J. Cell Biol. 151:639-652(2000).
CC -!- MISCELLANEOUS: The sequence shown here is derived from an
CC EMBL/GenBank/DDJ third party annotation (TPA) entry.
CC -!- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; BK000146; DAA00021.1; -;
DR HSP; O88792; IF97.
DR InterPro; IPR010939; DUF1136.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR011046; WD40-like.
DR Pfam; PF06582; DUF1136; 19.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD00066; SH3; 1.
DR PROSITE; PS50835; IG LIKE; 53.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
KW SEQUENCE 17903 AA; 2033688 MW; DSA06FF866E8B8B5B CRC64;

Query Match 4.0%; Score 204; DB 2; Length 17903;
Best Local Similarity 20.2%; Pred. No. 50;
Matches 244; Conservative 165; Mismatches 399; Indels 402; Gaps 63;

QY 7 TVALT-----LTVSVVTHNQVFSL-----VKEPILKQTQASSISGADYAESG-- 52
DB 10633 TVNLTSNLSNLSVSVVQVETSPVSLGSLATVBPQLKAMPVTKSNTLAYSEGGIP 10692
QY 53 -----KSKLKNETSGPDDVTVDLFSK-----RTTPEKIKDNLAKGR---E 93
DB 10693 VQVITVEAKLVNTRADVEENITVTEQPNWCIAASQLPAIEKKUENALQTPQFASE 10752
QY 94 QELKA---VTENTESEKQITSGSQLEQSKESL-----SLNKTVPSTSNWEICD-----FI 140
DB 10753 SILKTSPOISRNAHFETR-THREEYSTTTESLVTTQALRDDIDSTQKNVMDQVMYKHFA 10811
QY 141 TKGNTVLGSKSGVEKLSQTHLVLPQQAADGTQLIQVASFPTDPKKTATAEYTSRAGE 200
DB 10812 TK-----SLDKT---VKVETD-----TQATDTHIKQK-----TPTHK---LASSTQTITE 10852
QY 201 NGEISQLDVGKEELINEG-----EVPNSYLLKKVITPTGYKHIGODAFVDNKNIAEVN-- 253
DB 10853 SPEI-----IKTITISEGSGSKGMIRTLRIKKVK-----GNKQVTKIETVEEDDK 10900
QY 254 LPESLETISDYAFALHAKQI-DLPDNLKAIQELAFFDNQITGKLSLPRQLMLAERAFK 312
DB 10901 QPETTVTVEVPYEEKEPEEQELPEEVVV-ETVEDGPKKKKIRTVIKVKYKGDQKE 10959
QY 313 SNHINTIE---FRGNSLKVIGASQDNDLSQLM-LPDGLEKIESEAFGMPGDHYNR 368
DB 10960 VTKIETVEDDKQPEVTVTVEVPYEEKEPEEQELPEEVVVETVEDGPKKKKIRTR 11019
QY 369 VVLWTKSGKPSGLATENTYNPDKSLWQSPESPIDYTKWLEEDFTYQKNSVTGSNKGLO 428
DB 11020 VIKKVGKQKE---VTIETVEEDDK-----KAETTIVTEE-----TELSAPSVG 11061
QY 429 KYRKNKLEIPKQHGNGVITTEIGD-----NAFRNVDFQV-----462
DB 11062 KVLQKKRVIVQPEDAVTVFELPERKSVILSEKBDGTPTKTVIKTRIKKIQGPNMVEYTK 11121

QY 463 -KTLRKYDLSEVKLPSTIRKIGAF-----AFQS-----NNLKS 494
DB 11122 VQTVVEYE---KAPQIVSVKEKNTPPPELPEERLSEVVMLEPDEVSEAVDEBGRDKM 11177
QY 495 FE-----ASDLEBEIKE-GAFMNNRIE---TLSEKKOKLVITIG---DAAPHINHIYA 538
DB 11178 IKTKKRIIRKIPALDNTHEVTEIGIIEQDNVEPIYSVKIQERPLTESKPEBSKL----- 11230
QY 539 IVPESVQEIERSAFRQNGANNLIFMGSKVK--TLGEMAFLSNNLEHLD----- 585
DB 11231 IELPEHVTEL-----NVILPDGKKRRTVKSRAPKSLDDDLDEVTTIHIIEE 11279
QY 586 -----LSEKQKLTETIPQAFSDNA-----LKEVLLPASIKTTREEAPK--- 624
DB 11280 DKPELTKVNIWVPSDEISITPIEPLPEETVTEELDENKKPK-KTKTRTFKCRGP 11338
QY 625 --NHLKQLEVASALSHIAFNAL-----DNDQGD---EQFONKVVVVKTHNSYALAD---G 671
DB 11339 DDDEYFQIQTIDEGKEPISLIRVVSNDENADIIDISKLDDKVLKHKQPKHKDQYYK 11398
QY 672 EHFIVDPDKLSSTIVD-----LEKILKLEIG---LDYSTLRQTTQTFRDMT 715
DB 11399 EYTIPEEASADALQKPTKDKTKKQKKTLETPIEEVDETVIIDEGTGEQTDQI- 11452
QY 716 TAGKALLSKNLROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGQ----- 769
DB 11453 ---AIKRKPRKVQGNVQ-----VEAVD-EKPIEKKEKA-KKKKVVVKTKGDEMDDY 11497
QY 770 ---LLERSINKAVL-----AYNSAJK---KANVKRLEKELDLTLGLVEGKG 810
DB 11498 IHFLIHQEIPTVLQPYQRTMELPQRRDSSFKQPVKLTMPMKIEK-----VFEKK 11549
QY 811 P-----LAQATMVQGVYLLKTPLPPEYIYGLNVYFDKSGKLIYALDMSDTIGEGQKDAYG 866
DB 11550 PKMVEISSVVEFPQMLKAPKQRPQ-----SPKNKKLK--SWIRFVYPYCFPPVVVTEL 11579
QY 867 NPILNVDEBEGYHALAVATLADYEGLDIKTLNSKLSQTSIRQVPTAAY----- 917
DB 11580 -----KQNEA-----SPKNKKLK--SWIRFVYPYCFPPVVVTEL 11612
QY 918 ---HRAGIFQIAONAAAEOLLPKPGTHS---EKSSSES---ANSKDRGQSN-----P 964
DB 11613 ETNREVELSRNDEAEVLEKLRPKFKHSPKAELEADLGAYESDHSKSNKELLHP 11672
QY 965 KTNRGHSAI 974
DB 11673 KYRKGKKEKI 11682

RESULT 67

001385
ID 001385 PRELIMINARY; PRT; 2346 AA.
AC 001385;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Tpr homolog.
GN Name=Mtor;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97296455; PubMed=9152019;
RA Zimowska G., Aris J.P., Paddy M.R.;
RT "A Drosophila Tpr protein homolog is localized both in the
RT extrachromosomal channel network and to nuclear pore complexes.";
RL J. Cell Sci. 110:927-944(1997).
DR EMBL; U91980; AAC47506.1; -;
DR PIR; T13829; T13829.
DR FlyBase; FBgn0013756; Mtor.

DR GO: 0005643; C:nuclear pore; IDA.
DR GO: 0005634; C:nucleus; IDA.
SQ SEQUENCE 2346 AA; 262363 MW; B44FDA600B68934 CRC64;

Query Match
Best Local Similarity 4.0%; Score 203.5; DB 2; Length 2346;
Matches 224; Conservative 169; Mismatches 396; Indels 415; Gaps 44;

QY 6 KTVALLTL-----TTVSVVTHQEVFLSKEPIKQTOQASSSTISGADYAESGSKKL 56
DB 225 KTESLKLMEQVEQAVKTIIGELTSKIE---MNDTAFKQONQATEEVVGLKKELDAKEL 281
QY 57 KINETSQVDDTVTLDF---SDKRTTPEKIDNLAQGR-----EQELKAVTENTSEK 107
DB 282 -----FEIFKSTSDHLIQREELQGISIKRLLEAEAEQCAQLTEQMETWK 328
QY 108 QITSSQQLSQSLESLSNKTVPSTSNWEICDPTIKNTLVGLSKSGVEKLSQTDHLVLP 167
DB 329 QKHSALDEQNKKIQAEOELASAND-----LLKQARESNIESAICOLAP 373
QY 168 QAADGTQLIQVASFAPTPDKTAIABYTSRAGE---NGEISOLDVDGKEIINE----- 217
DB 374 SAASVRLIR-SDLST---ELYSWAKSSELEMRNCEISQKLQKLSIAEISGAP 428
QY 218 -----GEVFNYSLLKKVTIPTGYKHICQDAFVDNKNIAEYNLPESLETISDYAF 266
DB 429 ILEKQNSDYQKMEINSELLRE-----HDELLQNKLCLELERALSTLNHNQ 477
QY 267 AHLAKQI-----DLP-----DNKAIAGEAF 288
DB 478 ENKCLKQTHDLSRQVCMLLDELINCIRAGVKHVIQPTRLQPTSLSISDNLVTFSSIEE 537
QY 289 FDNQITKGLSLRQLMRLAERAFKS-----NHIKTIIFRGNLSKVGEASFQND 338
DB 538 LVDRNTYLLNKRLELTTELLEASEKQDKMLQSKNHIKRLDARFAELEDL---LTQKNT 595
QY 339 LSQLMLP-DGLEKIESEAFNPGDDHYNNRVLTWTKSGNPSGLATENTYVNDKSLWQ 397
DB 596 VTTLSKCDRYKLYFAA-----QKLGQNTVDDSDNLEFDSALDTS 639
QY 398 ESPEIDYTK--WLEEDFTYQKNSVTGFSNKGLOKVRNKNRILPQKHNGVTITEIGNAF 455
DB 640 EQPAANFEKSRKLEKVRVLEQOQLEG-----EVKKYASLK----- 674
QY 456 RNVDQFNKTLKYVD-LEEVKLSTIRKICAFQSNLKSFRASDDLEIKGAFWN--- 511
DB 675 ENYDYITSKRNDALAQFQDSMRKEVRLT---SSNCKLMNTTFQKEQIEILLHTNIGT 732
QY 512 --NRITELT-----LKDKLVITIGDAAFINHIYAIVLPESVQEIQR 550
DB 733 YKQOVTTLEERTYKYEKTIKHEQTVHLLKDEW-----AAHKHAAADAQAQSLRQENR 787
QY 551 -----SAFRQNGANNLIFMGSVKVTLGEMAFLSNRLE----- 582
DB 788 ILRDTSSRLQIEKETVHREQQSLL-----LMSLEFIKNTLSEMEGRQRLQORL 839
QY 583 -----HLDLSEOK---OLTEIPVQAFSDNALKEV-----LLPASLKITIREBA 621
DB 840 DDTVRELAQRHFQEEBEKPREFINEFKRQAFATKLDXDEKQADKQWQAEILTSVREL 899
QY 622 FKNHLKQLEVASALSHIAFNALDDNDGDEQFQDNKVVVVTNHSVALADGEHFIVDPKL 681
DB 900 AEKVN-KVNELSKLQEVLTPTLNDN-----HITANKARAEFEKLQDA 943
QY 682 SSTIVDLKILKLI--EGLDYSTRQTTQTQFRDMTTAGKALLSKSNLRFQGEKQKFLQRA 739
DB 944 TVVIESLTRELAKTRHGHQFYKMSQSAESEIKRLHELHGLVSK-----QEESIKKLRS 999
QY 740 QPFLGRVLDLDAJAKAEKALVTKATKNGQLLERSINKAVLAYNNSAIKKANVKRELEL 799
DB 1000 E-----AELKTRISDLAEAMLSNVTEQSKTNQS-----GQLKSAQ 1036
QY 800 DLLTGLVEGKGPLAQATMQGVYLLKTPLPPLPEYIIGLVNVPDKSGKLIYALDMSDTIGE 859

1037 DDLKSILLE-----KLTAEANTTIRLTSENTSLVESLNAAEV--- 1072
QY 860 QOKDAYGNPILNVDEDNEDGEGYHALAVATLADYEGLDIKILNLSKLSQTSIROVPTAAVHR 919
DB 1073 -----KYANGMTQ-----HSADIQELTRYKABFFKA--NDELNQLKSGRESLQAAY-- 1116
QY 920 AGIFQAIQNAAAAEAEQQLPKPGTHSEK----- 946
DB 1117 -----DELLRSNAEQAQKLLDKERSEKRVSDLHALNSNLHQIEALASKLAVLASQSNP 1172
QY 947 -SSSESANSKDRGLOSNPKT--NRGRHSAILPRTGSGSFYGLTYSVALLSLITAI 1003
DB 1173 NSSLSNESAMGDQSLNASGLTAAEGRNNE-----QLLKIKFL 1211
QY 1004 KKKK 1007
DB 1212 RKEK 1215

RESULT 68
QY Q8R6A0 PRELIMINARY; PRT; 921 AA.
AC Q8R6A0;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Exonuclease SBCC (EC 3.1.11.-).
GN OrderedLocustNames=FN0522;
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RX DOI=10.1128/JB.184.7.2005-2018.2002;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyripides N.C., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AB010564; AAL94718.1; -.
DR HSP; P58301; 1US8.
DR GO: 0016020; C:membrane; IEA.
DR GO: 0005634; C:nucleus; IEA.
DR GO: 0005524; F:ATP binding; IEA.
DR GO: 0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO: 0007059; P:chromosome segregation; IEA.
DR GO: 0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR Prodom; PD000006; ABC_transporter; 1.
KW Complete proteome.
SQ SEQUENCE 921 AA; 108848 MW; 3E4E21DF9EAF0CBE CRC64;

Query Match
Best Local Similarity 4.0%; Score 203; DB 2; Length 921;
Matches 183; Conservative 141; Mismatches 293; Indels 230; Gaps 43;

QY 49 ESSGSKLKNETSGPVDDTVTLDFSDKRTTPEKIDNLAQGRQELKAVTENTSEKQ 108
DB 56 KETGSIYKFEKSKVD-----IDFIANDGRYNNLKTFFKTKPKQ 98
QY 109 I---TSGSQL-----EQSKESLSNKTVPSTSNWEICDPTIKNTLVGLSKSGVEKLSQT 160
DB 99 TLKDMTSGEYDGIQEKLEELCGIKKGKPEETVENIV---IAKQNEFINIFKAKPNREI 155

```
QY 161 DHLVLPQAADGTQIOVASPAFTPDKKTATIAEYTSRAGE-NGEISQLDVGDKIINEGE 219
Db 156 FNKIP-----NTQIKEMYSFL---KEADVYKEKIKDLDKEITFL-----KENMEDKE 202
QY 220 VFNYSLLKKVITPTQYKHIGQDAFYDNKNIAEVLNIPESLETISDYAFALHAKLQIDLPDN 279
Db 203 QITNFKKEKELE---KNL-QDRP-KNINVUSKNLENE---IKDYETTEIELN--NLVKN 252
QY 280 LKAIGELAFFNQITGKLSLQPLMLRLAPKGS-----NHIKTIERFNGSLKIVIG 330
Db 253 IKD-----EENKIKYLNILNIEIAKQAKSKIIIVKETEKSYLEYLEIE-NRLKDLR 305
QY 331 EASFQNDLSQ-----LMLPDGLEKTESAFTGCPGDHNNRVVLTWTSKGNPSGLATEN 386
Db 306 EN--LDNLEBOKNIOYQNNIEKLE-----LSNKNLKNIDINLEEN 345
QY 387 TVYNPDKSLWQESPRIDYTKMLEEDFTY-----QKNSVTGFSNKGLO-KVKNKN 435
Db 346 ISKNSKKNLEES-RISNLKIKEEDLDLKKYISLLDELEKLENPKDKKLEDKLKTTE 404
QY 436 LEIPKQH-----NGVITEIGDNAPRVDFQNTKLAK-----YDLEEVKLPSTIR- 480
Db 405 IDILKELISKDLPTKNTINIEIEBKLSNFOLEKELKLLBEQKIIFIEIKTLKSSKE 464
QY 481 ---KIGAPAFQS--NNLKSFEASD-----DLEIEKEG-----AFMNNR 513
Db 465 LSDKICPFLNEKQCNLEDEADYPSKISIKEDLENLKNIEKTKQILVEKVVFEKK 524
QY 514 IETLEL-----KDKLVITIGDAAPHNHYAIV--LPESVQEIQRSAFRQNGANNLIFMGSK 567
Db 525 KQYPELEKSINKOLEISLKNEEINLKEIBLDIKNLDIDIOKLENGEFQNSQ-----MLRE 579
QY 568 VKTIGEMAF-----LSNRLHLDLSEKQKOLTEIPVQAFSDNALKVEL--LPASL 614
Db 580 KTELEVERNLDKRENKLNILENLEIESEKILK-----NONSIKSNLEEDVFS 632
QY 615 KTIREEAFK-----KNHLKQLEVASALSIAFNALDD--NDGDEQFDNKNVVKTHNSYA 667
Db 633 KKIKEDINKNIESIKSEIKTPE-----NKLDDLNKNPYNEVLKNVLAIE----- 675
QY 668 LADGHEFTVDPDKLSSTIVDL-----EKILKL---IEGLDYSTLRQTQTQOPDMTT 716
Db 676 --DLENLLIKVDNKNIKELYSLSRSDKNLKEKVFNLBEEKIKNIAELAEKYDTI----- 727
QY 717 AKGALLSKNLRQSKQKFLQEAQFPLGRVLDKAIKAEKALVTKKATKNGQLLERSIN 776
Db 728 --KEELNINKLSSQSEKIENYKILEKISSQE--EKQKLLVEFKK-----LENKFN 777
QY 777 KAVLAYN 783
Db 778 KASLIRN 784
```

RESULT 69

Q72X91

ID Q72X91 PRELIMINARY; PRT; 3242 AA.

AC Q72X91; 05-JUL-2004 (TREMBlrel. 27, Created)

DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)

DE LPXTG-motif cell wall anchor domain protein.

GN OrderedLocusNames=BCE5487;

OS Bacillus cereus (strain ATCC 10987).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=282523;

RN [1]

RP SEQUENCE FROM N.A.

RX PubMed=14960714; DOI=10.1093/nar/gkh258;

RA Raiko D.A., Ravel J., Oksas O.A., Helgason E., Cer R.Z., Jiang L.,

RA Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,

RA Nelson W.C., Kolatse A.-B., Fraser C.M., Read T.D.;

RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic

RT adaptations and a large plasmid related to Bacillus anthracis pXO1."

```
RL Nucleic Acids Res. 32:977-988 (2004).
CC -i- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (By similarity).
DR EMBL; AB017281; AAS44387.1; -.
DR TIGR; BCE5487; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005518; P:collagen binding; IEA.
DR InterPro; IPR008966; Adhes_bact.
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR001899; Gram_pos_bind.
DR Pfam; PF05738; Cna_B; 21.
DR Pfam; PF05737; Collagen_bind; 5.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRfams; TIGR01167; LPXTG_anchor; 1.
KW Cell wall, Complete proteome, Peptidoglycan-anchor.
SQ SEQUENCE 3242 AA; 358162 MW; 2AB79966B3245640 CRC64;
```

```
Query Match 4.0%; Score 203; DB 2; Length 3242;
Best Local Similarity 20.3%; Pred. No. 5.1;
Matches 238; Conservative 183; Mismatches 466; Indels 288; Gaps 63;
QY 17 VVTRNQEVFLVKEPI---LKQTQASSISGADYAESGSKL-KINETSGP----- 64
Db 2172 VETIAPHYDLNKKPIKPTIKKSTETLHITAKNALTKGGVELSKVDDIDGTLKLGAVN 2231
QY 65 -VDDTVTDLFSDKRTTTPKIKNLI-----AKGPREQELKAVTENTSEK-- 107
Db 2232 IVDMNGTVIHKDLVTS--KGNIEIDLRPDGYQIETKAPKHVNLNETPIHTIEKGQ 2288
QY 108 ----QITSGSQLESKESLSLNTKVPSTSNWEICD----FITKNTLV-----GLSKS 152
Db 2289 KKAISLTKAKNSLKGQ--SIELIKVDDLDQTKLADAVNLDQNGTQIKTDLKTNGEGKI 2346
QY 153 GVEKLSQTDHLVLPQAADGTQILQ--VASPAFTPDKKTATIAEYTSRAGENG-BISOL-D 208
Db 2347 TVENLRGTVQFVETIAPKHYDLDKPIVTVVEKSKQDAVTVTNSLTGKGVELTKVDD 2406
QY 209 VDGKEIINEGEVFNLSYLLKKVITPTGYKHIGQDAFYDNKNIAEVLN--LPESLETISDYAF 266
Db 2407 VDGTTL--EGAVFN-----IVDMNGTVIHKDLVTSKGIENLDRPGDYQIETKAP 2457
QY 267 AHLAKQIDLPDLNKAIGELAFFNQITGKLSLQPLMLRLAE-----RAPKSNHIKTI 320
Db 2458 KHYDLNKEPIPTTIEK-GQAE--PISVTAKNSLTGKAVELSKMDMDIDGTALGAIPKIVD 2514
QY 321 FRGNSLK-----VIGEASFQNDLSQLMLPDGLEKTESAFTGCPGDHNNRVVLTWK 374
Db 2515 MNGNDVETGTTDAKGVSIPTD-----LHPGDYQIETTA-----PKHY-----K 2554
QY 375 SGKNSPGLATENT-----YVNPDKSLWQESPEIDYTKMLEEDFTYQKNSVTGPNKGLQK 429
Db 2555 LDATPIKPTIEKSOAEKLVQTKAKNSLTGKAVELIKVDDINQDTKLSDAVFNIDAKG--K 2612
QY 430 VKRNKNLEIPQKHNGVITIEIGDNAPRVDFQ---NKTILKYDLEEVKLSITRIKIGAPA 486
Db 2613 VVRT-NLTTDKD-----GKVSANLREPQDYQFVETKAPKDYDLNKTPIPTTIEK----- 2660
QY 487 FQSNLKSFEASDLEBIKEGAFMNNRIETLEKDKLVITIGDAAPHNHYAIVLPSVSQ 546
Db 2661 SQTHV-SVTAKNGL---TKGGVELTKVDSLDAKE---TLEGAVFKIT----- 2701
QY 547 EIGRSAPRQNGANNLIFMGSKV--KTL--GEMAFLSNRL-EHLDLSQKQJTEIPVQAQS 601
Db 2702 DMNGNDIRTNLVTN---KDGKIIAKDLQPGDYQIETKAPKHVYDLNED-----PIKFTI 2752
QY 602 DNALKEVLLPASLKTIRSEAFKKNHKLQLEVASALSIAFNALDDNDGDRQFD-----N 655
Db 2753 ERSQTKHVFVTATNSLTGKSVELIKVDDVSENTTLEGAVFKIVNKGQHDVTRDTTQKNG 2812
QY 656 KVVV-----KTHNN-----SYALADGEHPIVD---PDKLSSTIVOLEKI 691
Db 2813 RLVDDELPGDYETKAPKHYDLNNETPIKFTVKKQEKIASVTATNSLTGKAVELSK-- 2871
```



```
Db 6023 NNAMTKEEIEQAQAQLA--QALQDIKDLVKAKEDAKNAIKAL-----AN 6064
Qy 931 ABAEQLLPKPGTHSEKSSSS--ESANSKDRGLQS--NPKT-----NRG 969
Db 6065 AKRDQINGNPDLTPBQKAKALKEIDEAEKRALQNVENAQTIDQLNRG 6111

RESULT 71
Q99U54 PRELIMINARY; PRT; 6713 AA.
AC Q99U54;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Bbha protein.
GN Name=ebha; OrderedLocusNames=SAI267;
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N315;
RX MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
RA Kuroda M., Ohba T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,
RA Kanahisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus".
RL Lancet 357:1225-1240 (2001).
DR EMBL; AP003133; BAB42527.1; -.
DR PIR; B89921; B89921.
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR011490; FIVAR.
DR InterPro; IPR02988; GA.
DR InterPro; IPR006530; YD.
DR Pfam; PF07564; DUF1542; 8.
DR Pfam; PF07554; FIVAR; 44.
DR Pfam; PF01468; GA; 46.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 1.
KW Complete proteome.
SQ SEQUENCE 6713 AA; 722339 MW; AF6BEDE226BE4888 CRC64;

Query Match 4.0%; Score 203; DB 2; Length 6713;
Best Local Similarity 19.5%; Pred. No. 14;
Matches 232; Conservative 183; Mismatches 396; Indels 376; Gaps 64;

Qy 34 KQTQASSISGAD-----YABSSGSKLKINETS GP-VDDTVTLPSDKRTTPBKIKON 86
Db 5050 QQTETAGSFNEDPKQDAYQAQVAKDLINQTNPTLDKSVQBLTQAVTT---AKDN 5106

Qy 87 L--AKGPREQELKAVTEN-----TESEKQITSGSQLEQSK 120
Db 5107 LHGDQKLARDQQAQVTTVNALPNLHNAQQUTDINAAPTETVQAQVHQVATLHDHAME 5166

Qy 121 SLSLNKTVPTSTNWBICDIPITKGNLTVLGSLSGVEKLSQTDHLVPSQAADGTQLIQVAS 180
Db 5167 TLK-NKV-----DQVNTDKAQPNTVTEASTDKKEAVD-----QALQAAQ 5203

Qy 181 FAFTPD-----KXTAIAEYTSRAGENGELSDVDGKEIINEGEVFNLSYLLKKVTIPTGY 235
Db 5204 SITDPTNGSNANKDAVEQALTKLQS--KVNEL--NGNERVABEAKTQAKQATIDQLT----- 5254

Qy 236 KHIGODAFVDNKN-----IAEV-----NLPESETTSDYAFALHAKQ-IDL--- 276
Db 5255 -HLNADQIATAKQNTDQATKQPIAEVLVDQATQLNCSMDQLQQAQVNEHANEVETIDYTQA 5313

Qy 277 -PDNLKAIGE-LAPFDN---QITGKLSLPRQLMRL--AERAP-----KSNHIKTIE 320
Db 5314 DSDKQKAYQAIAADAEVNLKQANQKQVQDQALQNLNAKQALNGDVERVALAKTNGKHID 5373
```

```
Qy 321 PRGNSLVIGEASF-----QNDLSQLM-----LPDGLKEKIESAFTCNPGDDHYN 367
Db 5374 -QLNALANNAQQDGFGRIDQSDNLNQIQIVDEAKALNRAMDQL-QSEITGNBGR----- 5426

Qy 368 RVLWTSGKNPSGLATENTVNPDKSLWQESP- IDYTKW-----LE 409
Db 5427 -----TKGSTN-----YVNADTVQVYDAVDKAKQALDKSSQNLTAQVIVKLN 5472

Qy 410 EDFTYQKNSVTGFS-----NKGQKVKR 432
Db 5473 DAVTAALKALNGEBRLNKRKAEALQRLDQLTHLNAQRLAIQIINNAETLNKASRAINR 5532

Qy 433 NKNLE-----IPQHNGV-----TITEIGDNAPRV-----DFQNKTLRK 467
Db 5533 ATKLDNAMGAVQVYIDEQHLGVISSTNYINADNLKANYDNAINAANAHELDKVGQNAIAK 5592

Qy 468 YDLEEVK-----LPSTIRKIGAFQSNL-----KSFEASDDLEE 503
Db 5593 ABAEQLLKQNIIDAQNALNGDQNLANAKDANAFVNSLNLNGLNQOQDILAHKAINNADTVSD 5652

Qy 504 IKEGAFMNNRIETLELKDCLVTIGDAAPHINHIYIVLPESVQBIHRSFQNGANNLIF 563
Db 5653 VTD--IVNNQID-----LNDAMETLKLVDNEIPNAEQTVNYQNAADDNAKTNFDD 5700

Qy 564 MGSVKVTL-----GEMAFLSNRLEHLDLSEQKQ-LTEIPVQAFSD---NALKE 607
Db 5701 AKRLANTLLNSDNTNVNDINGAIQVNDAIHNLNGDQRLQDAKDAKAIQSIQNALANTLKE 5760

Qy 608 VLLPASLKTREEAFKKNHLKQLEVASALSHIAFNALDNDGDQFD---NKVVVKTHN 664
Db 5761 I--EASNATDQDKLIANKKAEEL-ANSIINI--NKATSNQAVSQVQTAGNHAIEQVHAN 5815

Qy 665 SYALADGEHFIVDPDKLSSTIVDLK-ILKLEIGLDYS-----TLRQTTQTQ 710
Db 5816 EIPKAK-----IDANK-----DVDKQVQALIDIDRNPNTDKKQALKDRINQILQOG 5864

Qy 711 FRDMTTAGKALLSKSNLRQGEKQ--KFLQEAQFL-GRVDLDKAIAKAEKALVTK----- 762
Db 5865 HNDINNA-----LTKEEIEQAQAQAQALQDIKDLVKAKEDAKQDVQKQVQALIDEIDQNP 5920

Qy 763 -KATNGQOLLERSINKAVLAYNSAIAKKANVRLKELDLTLGLVEGKPLAQATMVQGV 821
Db 5921 NLTKQEKQALKDRINQ-ILQCHNGINNAMTK---EEIE-----QAKAQLAQA----- 5964

Qy 822 YLLKTPPLPEYIYGLNVYFDKSK- LIYALDMSDTTIGEGQKDAYGNPILNVDED--NEG 878
Db 5965 --LKEIKDLVKAKENAKQDVQKQVQALIDEIDQNPNTDKKQALKDRINQILQCHNDI 6022

Qy 879 YHAL-----AVATLADYBGL-DIKTILSKLSQLTSIROVPTAAVHRAGIFQAIQNA 930
Db 6023 NNAMTKEEIEQAQAQLA--QALQDIKDLVKAKEDAKNAIKAL-----AN 6064

Qy 931 ABAEQLLPKPGTHSEKSSSS--ESANSKDRGLQS--NPKT-----NRG 969
Db 6065 AKRDQINGNPDLTPBQKAKALKEIDEAEKRALQNVENAQTIDQLNRG 6111

RESULT 72
Q723X5 PRELIMINARY; PRT; 1775 AA.
AC Q723X5;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Cell wall surface anchor family protein.
GN OrderedLocusNames=LMOF2365_0350;
OS Listeria monocytogenes (serotype 4b / strain F2365).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=265669;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15115801; DOI=10.1093/nar/gkh562;
```



```
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF01463; LRRC1; 1.
DR Pfam: PF00560; LRRC1; 33.
DR PRINTS; PRO0019; LEURICHRPT.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
FT NON_TER 1 1316
FT NON_TER 1316 1316
SQ SEQUENCE 1316 AA; 145688 MW; C64B0D7172AF6336 CRC64;

Query Match
Best Local Similarity 4.0%; Score 201.5; DB 2; Length 1316;
Pred. No. 1.7;
Matches 240; Conservative 185; Mismatches 452; Indels 401; Gaps 59;

QY 26 SLVKEPILKQTOASSI-----SCADYAESGSKLKINETSPPVDT----- 68
DB 3 SILNAPVRKLPDSNNALSIGTKALFGVAEYLTELRLANLLGDSLNPISFTDVLQALQ 62

QY 69 ---VTDLFSKRTTPEKIKDNLAQPRE-----DELKAVTENTSEKQITSGSOLEQSK 119
DB 63 SLKVLDSGNRIIA---LEESIFDGNRLRVELYLQNKIATVPVAAVKELSLKLL--SL 117

QY 120 ELSLUNKTVPTSNW---EICDFITKGNLVLGSKGVEKLSQTDHLVL-----PS 167
DB 118 RSNRIESLEPDAPFANKLERDL--RNNRIRSLKSKAFANLASMKEVLLAGNQLSHIDE 175

QY 168 QAADGTOLIOVASPAFTPKTAIAYTSRAGENGESQLDVGKEIINEGEVFNYSLLK 227
DB 176 RALAGMDVLQKLDLS-----DNLLSEFPSEA--LGSVSVSLKVLNLSNIGKLESNHLAQ 228

QY 228 KYTIPTGYGHIGQDAFVNKNIAEYNLP-----ESLETISDYAFALHAL 271
DB 229 MKSL-----QIDISERTIATI--LPOTFREQLLLKYVLSLNSLTIEDDAFEGH-- 277

QY 272 KOIDLPDNLKAI---GELAFPDNOITGKL-----SLPRLQMLAERAPKSNHIKTE--- 320
DB 278 -----DNLQTEILRDNNILFTPGSALGRLPKLSNLYLDFNRVA--ALSSSLKSIOPEN 329

QY 321 -----FR-----GNSLKVYGEASPDND----- 338
DB 330 IRYLSLRNVIRELPDGPFAFRKUIYLDISGSLGVIQEDTFKLEGLTEILKLSFNRI 389

QY 339 ---LSQLMLPDGLEKTESEAFNCPGDD-----HYNNRVMLWTKGKNPSGLATENTYVN 390
DB 390 ASLRKFVLP-----KLRRDLDSANSIDDLAIDSFHSLHLLYLMNSGNEHVQVGT--RTMIY 444

QY 391 PKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVK-----RNKNLEIP--KQNG 444
DB 445 P-----LTKLQVIDLSHCGKMLQADLPNNNTDLRIALLNHNQ 482

QY 445 VTIISIG-----DNAP-----RNVDFQNTLRKYDLEVK--LPSTI 479
DB 483 LRVIEGTFLNLLNLFDSLAGNLEQLRPSRVNTVNLRLTLNRQNALQELRADMTTE 542

QY 480 KRIGAFAPQSNLKSFEASDDLE---EIKEGAFMNRIT-----LELKD--- 521
DB 543 TALEVLDASDNALKGF--ATNLKIIPRLKVLANNRLEVPPELIADLHLEVIDLAGN 601

QY 522 KLVITGDAAF-----HINHIYAIVLPSVQIGRSAPRONGANNLI--FMGSKVKTIGEMAF 576
DB 602 RLTTIQLDQFGRLANRLRY--LRANTIDSVDQMAFNHSSQLQILDAAANRLERSRF 659

QY 577 LSN-RLHLDLSEOK-----QLETPVQAFSD----- 602
DB 660 EGTLLDRDLSDNKLTLGLPEQVLAKSRVQLRSLVSLAGNKLASVPFAFGQDHDLTQL 719

QY 603 ----NALKEV-----LPLASL-----KTIREAPKKNHLKOLE 631
DB 720 DLSRNAIREVPASHMLMINAKHVDPSYNYPLTPEAIATVLGPKVTVRLINLACTGVREL 779

QY 632 VASA-----LSHIAFNALDNDGDQFDKNVVK-----THNSYALADGEHFIVDPDKL 681
DB 780 MLETPYLSLNSLNHNSIAV-----SAKAPEKVTLLERLDLSHNAIGDADG----- 825

QY 682 SSTIVDLKILKLEGLDYST--LRQTTQTOFRDMTTAGKALLSK-SNLROGEKQKPF--- 735
DB 826 ---LRDIWPKLALGYLDLSSNPRTITASTFERLDGLTELYIRELPETLRKKNAPKL 882

QY 736 ----LQBAOFF--LGRVDLDKAIKAERKALVTKATKNGQLLERSINKA-----VLAYN 783
DB 883 KDLTILQAYQPAKGYIDVQGIQVELPALAAVNDIEAKDNLESQOLQVLHHPKLTLLGLH 942

QY 784 NSAIKKANVRKLEKELDLLTGLVEGKGLAQATVWQVYLLKTP-----LPLP-BYYTGLN 838
DB 943 GYALQ-----SLSSGAPAGLRNKYLTVALHMTSLTALPPALLPLPRAHIDFS 991

QY 839 VYFQSKGKLI--YALDMSDTIGEGKDAYGNPILNVDEDNEGYHALAVAT--LADYEGLDI 895
DB 992 IAGSKVGTGLTPQFVSSLDLDRKNSIQIDGLATNPVRCDCQARFRRWILATKVQDVRCASP 1051

QY 896 KTIILNSKLSQ-----TSIROVPTAAHYRAGIFQAIONAA 930
DB 1052 PTVAGKLTVEGDSSELVDCGRSKATSTQPAASSTGSTRTPITISFVNYTSEVQMTNAA 1111

QY 931 ABASQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGHSAILP-----RTGSKGSFV 985
DB 1112 ATTEHDI-----IWSMPVPVSSSTTTRTKGSKSVS-----SAAMPPLKOMPSSANDDTLI 1161

QY 986 YGILGYTSVALLSLITAI 1003
DB 1162 IGIVGVVVVFILLIICI 1179

RESULT 74
Q869B8 PRELIMINARY; PRT; 1885 AA.
AC Q869B8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Kinesin-related protein K4.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OC NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98359834; PubMed=9693369;
RA de Hostos E.L., McCaffrey G., Sugang R., Pierce D.W., Vale R.D.;
RT "A developmentally regulated kinesin-related motor proteins from
RT Dictyostelium discoideum."
RL Mol. Biol. Cell 9:2093-2106(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Suyama E., Sutoh K.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB102780; BAC56912.1; -.
DR HSSP; P33173; 1161.
DR DictyBase; DB0191404; knd.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PRO0380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
DR ATP-binding; Microtubule; Motor protein.
SQ SEQUENCE 1885 AA; 218215 MW; 15686AED3B007EC7 CRC64;

Query Match
Best Local Similarity 4.0%; Score 201.5; DB 2; Length 1885;
Pred. No. 2.8;
Matches 213; Conservative 169; Mismatches 338; Indels 323; Gaps 54;

QY 21 NOEVFSL-----VKPEILKQTOASS-----SISGADYAESGSKLKINETS 64
DB 626 NOELLEIDQSKQSIQLNDKLLKLETKQSKQSIQLNLQLIDIESSSKNKKSF----- 679
```

```
QY 65 VDTVTDLFSKRTTPEKIKDNLAKPREQELKAVTENTSEKQI-----TSGSQLEQS-K 119
Db 680 --ENVLGVPEKSVRLAERLED-----KYFTKEIESKKQIETLANSYLQLETTVQ 726
QY 120 ELSLN-----KTVPTSTN-----W-----EICDFTKGNLTVG 148
Db 727 QQLNINQQSQQKIQSLNNNDIEQFLVWVLPKQVNGVFOENQMFQYIIELEBKNTLID 786
QY 149 LSKSGVEKLSQTDHVLVPSQADGTOLIQVAFPTPKKTAIAEYTSRAGENGESOLD 208
Db 787 LQKE-VEQNYLN--TLEQRNDQYQI-----EINQLTTEY-----NNQIQLE 827
QY 209 VDGKEIINEGEVFNFSYLLKKVT-----IPTGVKHIGQ-----DAP-VDN 246
Db 828 STNQKL--QTOLYN--LLANATQSTQTLQQLQTSQKEIDTLTNEIEQLKNQYDIIRDVN 883
QY 247 KNIAEVLNPESETISDYAFALHAKQIDLPNLKAGISLAFDQITKGLSLPQLMRL 306
Db 884 DNLSK-----ESLE-----LKQILLSKTQLEQLSLAQOQ-KGNIEIIOQL--- 924
QY 307 ARAFAKSNHKTIEFRGNLSKLVIGEASFQNDLSQLMLPDGLEKIESEAFNGPDHVN 366
Db 925 -ESIIVDNO-QSID-----QKIEFQSQODN-----OSIKQS-----YN 957
QY 367 NRVLWTKSGKPNPSGLATEN---TYVNDPKSLWQESPEIDYTKWLEEDFTYQKNSVTGF 422
Db 958 QLESTILTIAQSENQRLTENKQFITSLSNEIKSLFNSIQOQKETIQLEFEFQEKQFDSL 1017
QY 423 SNKGLOKVRKNRNKLEIPKQHNGVTITEIGDNAPRVDFONKTRKYVDLEEVLPSTIRKI 482
Db 1018 LTNYNQLFKYNDLATSNESNRLEFDQFKDS-----NQSTQSLSELSLSK----- 1065
QY 483 GAFQFQSNLKSFEASDDLEEKEGAFQNRLETLE--LKDKLVITIGDAFHNHYAIV 540
Db 1066 -----ENDNL-----LQSSLLKSQLESIEKQKODQLIPI-----QLE 1098
QY 541 LPESVQEIQR--SAFRQNGANNLIFMGSKVTKLTGEMAPLSNRLEHLDSEQQLTEIPQ 598
Db 1099 LESKCKELSKLSQSEQTKQ---VTQLLISVDQYKISTNKLESQISDRNEINNKLK 1154
QY 599 AFSQNALKEVLLPASLK-----TIREAFKQNLKQLEVASALSHPALD 645
Db 1155 ATEINALKEENI--SLKQQLTKLKKAPKQTDREKDMIKKELEKLE-----KFNAID 1205
QY 646 DNDGDEQFQNVKVVKTHNSYALADGEHEIIVDPKL--STIVDLEKILKL-----LEGD 699
Db 1206 AK-LKQAIQDKQTIQSEKQSL-----EREIKDKRSHTSTETELDKLKKTHLAADVSKD 1259
QY 700 YSTLRQT-----TOTQPRDMTTAGKALLSKNLQGEKQKQFLOEAPFLGRVLDLKA 753
Db 1260 FIALNKSVEILTKSQELKSTIELESDSLKKNIELEKQEL-----VTLNQ--D 1308
QY 754 KAEKALVTKKATNGQLLERSINKAVLAYNNSAIKCAN---VKRLEKE-LDLTLGLVEG 808
Db 1309 KLEK-----EKKTNQLESDDHSATIKLENVENQITOLTSEIIDLSKQFQBF 1354
QY 809 KGFLAQATWQGVYLLKT-----PLPIPEYIGLVNVPFDSKGL--IYALDMSDT 856
Db 1355 KSSESNTKQBEINLKESNDLNQQLTNDNLKLTQSLDKLVFDFKSKQLMSTRSESNDT 1414
QY 857 IGEQKQDAYGNPILNVDENEGHALAVATLADYEGLDIKT-----ILNSKLSQTSIRQV 912
Db 1415 IKELQES-----IISKDKERQ-LTSEQVLKLTQD--INLKTWEYNDLNSQCQQLT----- 1461
QY 913 PTAAYHRAGIFOAIONAAAEQ 935
Db 1462 -----KTLQNVKSSNEQ 1473
PRELIMINARY; PRT; 1578 AA.
RESULT 75
Q9AV25
ID Q9AV25
```

```

QY 390 NPD----KSLWQSPRI-DYTKWLEBEDFYQKNS-----VTGFSNKGLOKVK 431
Db 772 NLEHISKLEHENVLSBFIQSGLESQLYLANEKELSMQMDERSLITNLKDB-LEQVE 830
QY 432 RNK-NLEIPKQHGVTITEIGNAFRNDVPQNTLR-----KYDLREVKLPS 477
Db 831 AQVKELQMDERSLITNLKDB-LEQVEAQVKELQMDERSLITNLKDBLEQVE--- 886
QY 478 TIRKIGAFQSNLKSFP-----EASDDLEIEKEG-----APMNRRIE-----TLELK 520
Db 887 -AQKV---ELKENQLESRRHLSVEQDSEALRRSNAKLQATVDHVVECKSLQTUTADLK 942
QY 521 DKLVTIGDAFHINHIYALVPSVQETGRSAFRONGANNLIFMGSKVKTLQEM-AFLSN 579
Db 943 KQKLEV-----HGYASHLEQELQSKRTM-----DFCKTLESLEAKLS 982
QY 580 RLEHLDLGEOKLTPIPVQAFSDNALKEVLLPASLKTIREAFKNHL-----KOLE 631
Db 983 LQEDISLKEQSLSLSL-----ENIFQE-----HKEHEERIDRVHLLNKKIEKKTVE 1029
QY 632 VASALSHIAFNALDNDGDEQFDNKKVVTTHNSYALADGEHPIVDPDKLSSTIVDLR-- 689
Db 1030 LSNLEREVSILTAQLSSTEEBRESSTLDTIREVSILRADKAKLEANLDVNAQMIHYESQ 1089
QY 690 -----KILKLEGL-----DYSTLRQTQTOFRMTTAGKALLS---KS 725
Db 1090 LEDLRBSKTKIKDVLDSLNASKQNEMLTTDDNMRRSIEAARSNEDNLRKTLCLBELKS 1149
QY 726 NLRQKEKQKF-----LQEAQFFLGRVD-LDKATA----- 753
Db 1150 KSSDYEQKQIIBEISVLKIQVHIAQLODEVULTLOSSLDBAKFEKGKLEGLIQSUSECE 1209
QY 754 ---KABKALVTK-----ATKQGLLERSINKAVLAYNNSAIKXANVRLEKELDLL 802
Db 1210 ELKAKQKMLTDKVSQMDTLNANEGKQIEISAQTKLVMLGDEPPVKETSDVLEAKLSE 1269
QY 803 TGLVEGKPLAQATWQVGYLLKTPLEPIYIGLVNVTGSKGIYALDMSDTIGEQK 862
Db 1270 LSIIRG-----ANSEYQKIIYSIQEE---NEDLTRENQLMEKE-----LDLKS 1310
QY 863 DAYGNPILNVDNDEGYHALAVATLADYEGDLIKTILNKSQLSIRQVPTAAVHRAGI 922
Db 1311 -----QNKDENTKQVSLQDEVULMQSSLDEAFKENGKLEGL----- 1347
QY 923 FOAIQNAABEQLLPKPGTHSEK-SSSESANSKDRGLQ 961
Db 1348 ---LQSLSECEBELKAKQKMLTDKVSQMDTLNANEGKQ 1384

RESULT 76
Q7MRQ2
ID Q7MRQ2 PRELIMINARY; PRT; 2133 AA.
AC Q7MRQ2;
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE FmtB-like protein (Fragment).
GN Name=mxp-homologue;
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxId=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sweden309, Glasgow3759, LIM2, LIM3, LIM1, Southampton23,
RC Sweden307, Norway1018, FranceDuf, LiverpoolAG, and Glasgow3700;
RA Wootton M., Avison M.B., Bennett P.M., Howe R.A., MacGowan A.P.,
RA Walsh T.R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ567427; CAD98837.1; -
DR EMBL; AJ567429; CAD98839.1; -
DR EMBL; AJ567430; CAD98840.1; -
DR EMBL; AJ567431; CAD98841.1; -

```

```

DR EMBL; AJ567432; CAD98842.1; -
DR EMBL; AJ567433; CAD98843.1; -
DR EMBL; AJ567434; CAD98844.1; -
DR EMBL; AJ567435; CAD98845.1; -
DR EMBL; AJ567436; CAD98846.1; -
DR EMBL; AJ567437; CAD98847.1; -
DR EMBL; AJ567438; CAD98848.1; -
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR011490; FIVAR.
DR Pfam; PF07564; DUF1542; 17.
DR Pfam; PF07554; FIVAR; 1.
FT NON_TER 1
FT NON_TER 2133
SQ SEQUENCE 2133 AA; 232403 MW; B4F33BE54FF0A896 CRC64;

Query Match 4.0%; Score 201; DB 2; Length 2133;
Best Local Similarity 19.2%; Pred. No. 3.5;
Matches 217; Conservative 183; Mismatches 366; Indels 342; Gaps 58;

QY 20 HNOVFSLVKEPIKQTQASSISGADYAESGSKLKINETSGPVDVTVDLFSD---- 75
Db 1177 YNAKLAEINATPDATNDKNAINTLNQDRQQAIESIKQANTNAEVQQAATVAENND 1236
QY 76 -----KRTTPEKIKDLAKGPREQELKAVTENTSESKQITSGSOLEQSKB--- 125
Db 1237 QDVVVKQAARDKITAFAVAK--RIEAVKQTPNATDEBKQ-AAVQINQLKQQAQINQIN 1293
QY 126 KT---VPSTSNWEI-----CDFITKNTLVGLSGVKEKLSQTDHLVLPQAAQDGTQL 175
Db 1294 QTNQDQVTTTQAVNAIDNVEAEVVVKPAIADIEKAVKEKQKQIDNSL---DSTDNEK- 1349
QY 176 IQVASPAPTDDKTAIABYTSRAGENGESISOLDVDGKEIINEGEVFNFSYLLKKVT--IPT 233
Db 1350 -EVASQALAKEKEKALAA-IDQAQTSQVQAATNGVSAIK-----IIQPTKVKPA 1399
QY 234 GYKHIGDQAFVDNKNIAEAVNLPE-----SLETISDYAFALAL-----KQID 275
Db 1400 AREKINOKA---NEIRAKINQDKEATABERQVALDKINE--FVQAMTDTTNNRTNQVD 1454
QY 276 LPDNLKAIQELAPFDNQITGKLSLPQMLRAER-----AFKSNHKTITIEFGNSLKV 328
Db 1455 -DTTSQALDSIA-----LVAPHEIVRAAADARDAVKQYEAQKQIEQAHAHATDEBKQ 1504
QY 329 IGEASFQDND-----LSQLMPLDGLKIESEATGPGDDHYNVNRVVLVTGSKNPSGL 382
Db 1505 VALNQLANNEKALQINQAVTNDVKRVETNGIATLKG---VQPHIVI-----KPEAQQ 1556
QY 383 ATENTYVNPDKSLWOESP-----EIDYTKWL-----EEDFTYQKNSVTGFSN--- 424
Db 1557 AIKATAENQVESI-KDTPHATVDELDEANQLISDTLKAQQEIENTNQDAAVTDVRNGTI 1615
QY 425 KGLQKVK-----RNKN-----LEIPKQHGVTITEIGD--NAFRNVD 459
Db 1616 KAIEQIKPKVRRKRAALDSIENNNKQLDAIRNTLDTTQDERDVAIDLANKIVTKINDI 1675
QY 460 FQNTLRYDLER-----VKL-----PSTIRKIGAFAPQSNLKSPEASD-DLEEI 504
Db 1676 AQNTKNAEVDRTETDGDNDNIKVILPKVQVPAARQSVGVKABQAQNAL--IDQSDLSTEE 1733
QY 505 KEGA-----PMNNRIETLELKDGLVTIGDAAFHINHIYALVLP-----ESVQBIG 549
Db 1734 RLAAKHLVQALNQAIQDINHADTAQVQNQDSINAQNIISKIKPATTVKATALQIQNIA 1793
QY 550 RSAFRQNGANN-----LIPMGSKVK---TLGBMAFL-----SNRLEHLD 585
Db 1794 TNKINLIKANNEATDEEQNIAIAQVEKELIKAKQIASAVTNADVAVLLHDKKEIREIE 1853
QY 586 -----LSEKQLTEIPVQAFSDNALKEVLLPASLKTIREAFKKNHLKOLEVASALSHI 639
Db 1854 PVINRKASAREQLTTL-----FND---KQQAIEANIQTAVES--RNSLTAQLQ---NIYDT 1901
QY 640 APNALDNDGDGEQFDN--KVVKVTHNSYALADGEHPIVDPDKLSSTIVDLKILKLEIG 697

```

Db 1902 ATGQIDQDRSNAQVDTASINLTQTHDL-----DVHPKPKDAEKTINDILARVLTALVQ- 1955
 QY 698 LDYSTLRQTQTQFRDMTGTAGKALLSKSNLRQGEKQKFLQEAQFGLGRVDLQKALAKAEK 757
 Db 1956 -----NYRKVSRENK-----ADALK 1970
 QY 758 ALVTKATKNGQLLERSIN---KAVLAYNNSAIKXANVRLKELDLGLVGEKGLAQ 814
 Db 1971 AITALKLQWDEELKARTNADVAVLKRFNVALSDIEAVITEKENSLLR--IDN---IAQ 2025
 QY 815 ATWQGVLLKTPPLPEYVIGLVYFDSKGLIYALDMSDTIGBQKDAYGNPILNVDE 874
 Db 2026 QT-----YAKFAIATPEQLAKVKVLIDQ-----YVAD-----GNRM--IDE 2060
 QY 875 DNEGHALAVATLADYEGDIDKILNSKLSQLTSTSRQVPTAAYHRAGIFQAIONAAAEAE 934
 Db 2061 D-----ATLN-----DIKHQTFIVDEILAIK-LPAEA-----TKVSPKEI 2095
 QY 935 QLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGHRSAILPRTGSKG 982
 Db 2096 QPAPKVCTPIKKEETHESKRKEVE-----LNTGSEG 2127

RESULT 77
 Q7RNW8
 ID Q7RNW8 PRELIMINARY; PRT; 2911 AA.
 AC Q7RNW8;
 DT 01-MAR-2004 (TREMBlrel. 26, Created)
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Acetyl-CoA carboxylase 1 precursor-related.
 GN Name=PY01695;
 OS Plasmodium Yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxId=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=1236865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M., Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L., Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L., Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V., Cho J.K., Quackenbush J., Sedegah M., Shoaihi A., Cummings L.M., Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A., Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R., Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J., Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii.";
 RL Nature 419:512-519 (2002).
 CC -! CAUTION: the sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
 CC EMBL: AABL01000457; BAA21055.1; --
 DR HSSP; P24182; 1BNC.
 DR GO; GO:0003343; C:biotin carboxylase complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004075; F:biotin carboxylase activity; IEA.
 DR GO; GO:0016874; F:ligase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR005482; Biotin carb C.
 DR InterPro; IPR000089; Biotin lipoyl.
 DR InterPro; IPR000022; Carboxyl trans.
 DR InterPro; IPR005481; CPase L.N.
 DR InterPro; IPR005479; Cphb synth L_D2.
 DR Pfam; PF02785; Biotin carb C. 1.
 DR Pfam; PF03164; Biotin lipoyl; 1.
 DR Pfam; PF01039; Carboxyl trans; 1.
 DR Pfam; PF00289; CPase L chain; 1.
 DR Pfam; PF02786; CPase L_D2; 1.
 DR PROSITE; PS00866; CPASE_1; 1.

DR PROSITE; PS00867; CPASE 2; UNKNOWN 1.
 SQ SEQUENCE 2911 AA; 337275 MW; P2A9083552BD9E02 CRC64;
 Query Match 4.0%; Score 201; DB 2; Length 2911;
 Best Local Similarity 19.0%; Pred. No. 5.4; Mismatches 364; Indels 286; Gaps 44;
 Matches 192; Conservative 167;
 QY 1 MKKHLKTLVALTUT-TVSVV--THNQEVFSLVKPELILKQTOASSS----- 41
 Db 1382 VKKYLTLNLSLTNTINKMSTTPNEELINDISENNLSLTNTMHNDSIVNSGNSDSIVSOR 1441
 QY 42 -----ISGADYAESGSKLKLINETSVPVDDVTDLFSDKRTTPPKI----- 83
 Db 1442 TNNISEEKIDNNKLLISYDNVKNQYKDKIVSSSEIDFPVNT-----KDKLDRKMLISD 1496
 QY 84 -----KDLAKGPREQELKAVTENTESEKQITSSQLEQSKESLSLNKTVSTSNWEIC 137
 Db 1497 TEMSRSDNDSOSSNR-----TNTNNEKIELNFDNKKOKE-----NEFFLELSNMKRI 1546
 QY 138 DPITKGNLTAVGLSKSGVEKLSQTDHLVLPSSQAADGTQLIQVASFAFTPDKTAIAEYTSR 197
 Db 1547 EYLLKGYX-----QDYKCFED-----FLKDXNNFDKNLNYI 1578
 QY 198 AGENGEISQDVGKEIINEGEVFNLSYLLKKVTIPTGYKHIGQDAFVNKNIAEVLNPS 257
 Db 1579 LDKFTIYNHL-FSKKEIISEVEIY--YIL-----YKSIS-----DKK-----KQ 1614
 QY 258 LETTSDYAFALHALKQIDLDPN-----LKAIGELAFPDNQITQK-LSLPRQ 302
 Db 1615 YEIHAYTNYLSIKLILNKLNNLINSKQTNLSLOILNKIKILKEFKGIFGSIILLRY 1674
 QY 303 LMLRAERAFKSNHIKTIEFRGNS-----LKVIGEASFQDN-DLSQMLPDGL 348
 Db 1675 IILYEQNEKIHLYES-SFQYNKQMEDKKQSLQNYIRFLTSSWNNDYKIDQILNNTY 1733
 QY 349 EKIESEAFTPGDDHNNRVVLMTSGKNPSGLATENTYV-----NPDKS 394
 Db 1734 EE-KIKFLKSPSDIHYSLPLFKT---ENVYQEMFNTYLNHLYKYSTIKAKHSPNCL 1789
 QY 395 LW-----QESPEIDYTKLEEDFTYK-----NSVTGF 422
 Db 1790 VFSINNHYNNVMLNLEDITIEQPKSMLDDILYEMKSDKINTFHIINKNETFFNSLQNY 1849
 QY 423 SNKGLQKVKR-----NKLEIPKQHGVTITIGIGNAFNVDVFNQNTLKRKYDLEEVK 474
 Db 1850 LDNAFKKVTYLYINYFNKGYGEIHK-----INNNAKEREKIQDNLEKSIETNK 1900
 QY 475 LPSTI-RKIGAFQSNLKSFEASDDLE-----EIKEGAFMNNRIETLEKD 521
 Db 1901 NDSTINNKIYLLPFEEILTSNTASNFELKNVQGVFNKTKILGLTYKNNKYTSLPAQ 1960
 QY 522 KLVITGDAAFHINHIYALVLPESVQEGISAFRQNGANNLIPMSGKVTGLGEMAFSLNRL 581
 Db 1961 RIIDINDLQKQVDSNLIKVDQ-ENVGMT---KNGKNNILL--SKNIENGQM---NNQI 2011
 QY 582 EHLDLSEOKOLTE-----IPVAFSDNALKEVLLPASLKTIRE-----BAFKK- 624
 Db 2012 YDIFMSELKESIEDISIVRLNPSRDIKISNIYHIVLPNN-SNIEQINGIKIYKNF 2070
 QY 625 -NHLKQLEVASALSHIAFNALDDNDGDFQFNKVVVVKTHNSYALADGEHFIVDP-DKLS 682
 Db 2071 LNKYNELFIENYNNVYIKVYKEIEEKNKNETKLELQKLLHLVFDNEKEVIEKVDKIN 2130
 QY 683 STIVDLKILKILBGLDYSTLRQTTQTFQDRMTAGKALLSKNSURQGEKQKFLQEAQFF 742
 Db 2131 PVIID-ALYLLKRGARVDVTI-----YAYDFVKLINIALARNINTNDNFYINYSVKEF 2183
 QY 743 LGRVDLDK-----AJAKAEKALVTKKATNGQLLERSINKAVLAYNNSAIKXANV 792
 Db 2184 --KIDYSENSEENDTVSVGKESQPNKNDKQSKWLFENFDNL-----SDE 2230
 QY 793 KRLKEKELDLATGLVEGKGLAQATWVQGVYLLKTPPLPEYVIGLVNLYF 841

```
Db 2231 IRIKSLYLSNSLEIGQKMSVIGLLMNV---KT-----DEYKEGREDIAF 2272

RESULT 78
ID Q7YWF0 PRELIMINARY; PRT; 3080 AA.
AC Q7YWF0;
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Normocyte binding protein 2a.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxid=5933;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVO;
RX MEDLINE=20402589; PubMed=10920203; DOI=10.1073/pnas.160469097;
RA Rayner J.C., Galinski M.R., Ingravallo P., Barnwell J.W.;
RT "Two Plasmodium falciparum genes express merozoite proteins that are
related to Plasmodium vivax and Plasmodium yoelii adhesive proteins
involved in host cell selection and invasion.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:9648-9653(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVO;
RA Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,
RA Corredor V.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV138499; AAN39446.1; -;
SQ SEQUENCE 3080 AA, 363854 MW, 73PBD07DC5648968 CRC64;

Query Match 4.0%; Score 201; DB 2; Length 3080;
Best Local Similarity 19.7%; Pred. No. 5.8;
Matches 231; Conservative 189; Mismatches 395; Indels 356; Gaps 59;

Qy 32 ILKQTQASSISGADYASSGSKLKINETS-----GPVDDTVTLPSDKETTEKIKDN 86
Db 2025 LVKCKEUREISTALY-----DKIQTTSVINKRENDISNNI-DIVSKNLNEIDAIOYN 2076
Qy 87 LAKGPREQELKAVTENTSEKQI--TSGSQLEQSKSKE--SLSLKNKTPVSTSNWEICDPIT 141
Db 2077 FEK-----YKEIFDNVEYKTLDDTKNAYIVKKAELKQVDINKTKE-----DLDI 2122
Qy 142 KGNLTVLGSKGVEKSLQTDHLVLPQAAADGTQLIQVAFPTDPKTAIAEYTSRAGEN 201
Db 2123 YFNDLDELEK-----LTLSSNEMETKIVQ-----NSYNSFSFDIN 2158
Qy 202 GEISQLDYDGK-----EINNEGEVFN-----SYLLKKVTIPTG-----234
Db 2159 KNINDIDKEMKTLIPMLDELLNEGHNDISLYNFIRNIQIKIGNDKINREQENDTWIC 2218
Qy 235 -----YKHIGQDAFVNK-----NIAEYNLPESLETISDYAPAHALAKQI 274
Db 2219 FEYIQNNYFNKSDISIFNKYDDHKVDNYISNNIDVVKNSL--LSEH-----VINAT 2271
Qy 275 DLPDN-LKAIGELAFDNOITGKLSLR--QLMRLAERAPK-----SNHIKTIEFR---322
Db 2272 NIENIMTSIVEI-----NEDTEMNLEETQDKLELYENFKKKNIIINNKKIVHFNK 2327
Qy 323 --GRSLKVIGEASFOQDNLSQLMPLDGLKTESEAFPTGDDHYNRNVLTGSKGKPS 380
Db 2328 EINSLETYSISTNFNKNINETQNDILK-----NEFNN---IKTKINDKVK 2371
Qy 381 GLATENTYVNPDKSLQWQSPET-----DYTKWLEEDFTYQKNSVTGFSNKGLOKVKR--432
Db 2372 EL-----VHVDSTLTLSIQTFNNLYGLDMSNIQDVYKE-----DINNVELKKVKLYI 2420
Qy 433 -----NKNLEIPKQHGNG--VTITEIGDGNAPRNVDFQNK 463
Db 2421 ENITNLLGRINTFIKELDKYQDNNGIDKITYEINKENNSYIILKEKANNLKENF---SK 2477
Qy 464 TLKRYDLEEVKLPSITRKIGAFAPQSNLKSFEASDDLEEKKEGAFMNRIE-----TLEL 519
```



```
QY 282 AIGLAFD-----NQITGKLSLPLQMLRAERAFKSNH--IKTIEFRGNSL 326
Db 783 KI-----FDEINQKYDDITQNVNQNDKIVLRQM-----ENLKKYBEIKKENIQDRDI 833
QY 327 KVIQASFPQNDLSQLMDGLKIESEAFNPGDDHNNRNVVLWLTGSKGNPSGLATEN 386
Db 834 KYIGE-----NNDGNNNNNDNII--NGNNHTG----- 859
QY 387 TVNPDKSLMQBSPIDYT-----KWLBEPTYQNSV-TGFSNKGLOKVRNKNLIP 439
Db 860-----DYNVTLFHYTDLWNSQFTRTKENIQTNINN--IKSLIRNLQNEL- 902
QY 440 KOHNGVTITEIGDNAP-----RNVDFONKTLRKYLDEEV-----KLPSTIR 480
Db 903--NNTNTLKSNGSHFYDKIHTLEELKILTQEDINDKNVRKIIDIENTYQNDLHNEEIK 960
QY 481 KIGAFAPSNLKSFEASDLEIEKEGAFMNNRIETLEKOKLVIGDAAAPHINHIYAI- 539
Db 961 NITSYIKINILNIIIS-----IKQTYNNKNSISLKLKINNLT-NSTQEVINQKIP 1014
QY 540--VLPSVQBIGRS-----AFRONGANNL-----IFMGSKVKTIGEMAPL 577
Db 1015 TNLPLPEHIKQKSLSELNIMYKQIDKLNENVINLTYKSKDSLOFYINEKNKN-----I 1068
QY 578 SNRLEH-----LDLSEKQLTETIPVQAFSDNALKEVLLPA-----SLKTIREBAPK--- 623
Db 1069 NNNNDHNVNDYNDIKEN-----QIYKNNKLYECIQIKKOIDEIYNDYDQLFKNIS 1120
QY 624-----KNHL-----KOLEVASALSHI----- 639
Db 1121 QVNNHNSLSFTHSLNNHMSLIPQVTKYGHKNQOILSDIDNVIKQNEHKBSTYNLDTNSI 1180
QY 640-----AFNALDON--DGDEQDNKVVVTHHNSVALAGEHPIVDP----- 678
Db 1181 QLIKEIQKYFLQIPHILQENITTFENQYKDLIIKNNHKINNKLKDIITHIVINDYTLQEQ 1240
QY 679----DKLASTIVDL-----EKILKLLEGD-----Y 700
Db 1241 NNINYELQNKIKQIKRVNEVFTNNINYOQILNYSQADNFPNIPMKFQNNINDINRKY 1300
QY 701 SLRTQTTQ-----TOFRDMTTAGKALLSKNLROGEKQKFLQEAQFFLGRVDL 748
Db 1301 NVQKITEIINSYDIINYNKNNIKDMYQQLKNIQOQLNTTETQLNHIKQINHF----- 1354
QY 749 DKAIAKASKALVTKATKNGQLLERSINKAVLAYNNSAIKANVRLEKELDLTLGLVEG 808
Db 1355--KYFYKSNQTSIVKNIQNEKLKIQEFNKKIQHPK-----BQTQIMINKLIQ- 1400
QY 809 KGPLAQATWQGVYLLKTPPLPEYVIGLVYFDKSGKLIYALDMSDTIGEGQKD---AY 865
Db 1401-----PSNIHLHKMLPITQOQANTHLRNEHTK-----NDTKSYNMKDEENAM 1444
QY 866 GNPILNVDEBNEGYHALAVATLADYEGL-----DIKTILNSKLS 904
Db 1445 GYGTIN--ERKNSGTNDMINSITGDNTVLTNDQBERKDTSRNNNIHTDEKINNHETN 1503
QY 905 QLTSTRQVPTAAHYRAGIPQAIQNAABEQLL-----PKPGTHSKS 947
Db 1504 EQYHPKEHPNHYHQNDKESILQHTKINTTSORTIDSNMDRNNRYNTSSQKNNLHTNN 1563
QY 948 SSSSESANSKDRGLQSNPKNTGRHSAILPRTGSKGSF--VYGILGYTSVALLSLITA 1002
Db 1564 SNSRYNNNHDK--QNEHKYNOQKSS-----GKNSYWRIFVYAGGITAVLLCSSTA 1612
```

RESULT 81

Q6BU09

ID Q6BU09 PRELIMINARY; PRT; 2042 AA.

AC Q6BU09

DT 25-OCT-2004 (TremBLrel. 28, Created)

DT 25-OCT-2004 (TremBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TremBLrel. 28, Last annotation update)

DE Similar to sp|P25386 Saccharomyces cerevisiae YDL058w US01.

```
GN ORFNames=DEHA0C09658g;
OS Debaryomyces hanseii CBS767.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=284592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genolevures;
RA DuJon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., March C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisarame A., Boyer J., Catolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi R.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozler-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Winkler P., Souciet J.L.;
RT *Genome evolution in yeasts.*;
RL Nature 430:35-44 (2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genoscope;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, CR382135; CAG86127.1; -
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008565; P:protein transporter activity; IEA.
DR GO; GO:0006886; P:intracellular protein transport; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000357; HEAT.
DR InterPro; IPR003900; KID_repeat.
DR InterPro; IPR006955; Uso1_p115_C.
DR InterPro; IPR006953; Uso1_p115_head.
DR Pfam; PF02985; HEAT; 1.
DR Pfam; PF02524; KID; 3.
DR Pfam; PF04871; Uso1_p115_C; 1.
DR Pfam; PF04863; Uso1_p115_head; 1.
DR Pfam; PF04863; Uso1_p115_head; 1.
SQ SEQUENCE 2042 AA; 233328 MW; 959292DDEBEE1158 CRC64;

Query Match 3.9%; Score 200; DB 2; Length 2042;
Best Local Similarity 20.4%; Pred. No. 3.6;
Matches 189; Conservative 150; Mismatches 302; Indels 286; Gaps 43;

QY 20 HNOVFSLVKEPI-----LKQTQASSISGADYAESSGSKL-----KINFTSGPVD 66
Db 1161 YNDAIEKLNKNDISIASLKEH-SKKVSELD---SGHSLSQDLEAANSCLTEKQIK 1215
QY 67 DTV--TDLFSDKRTTPKIKONLAKGPREQELKAVTENSEKQITSGSQLESLSL 124
Db 1216 EHLSSNLADQISALEKVGK-----ELEASINNAEQE---SNKSRFEKEKAE 1263
QY 125 NKTVPSTSNWEICDFITKGNLTVLGSKSGVEK---LSQTDHLVLPQAAQTQIQVASF 181
Db 1264 NQNL-----TNLEAKQKAEKRLDLVQEEKAEKELAKLQILDNSK 1307
QY 182 APTDPKTAIABYTSRAGENGESIQLDVGDGKEIINEGEVFNFSYLLKKVITPTGYKHIGQD 241
Db 1308 LET-----EVSELKSDITKPKDEHTIIN---EKLSTKELSEKKD 1345
QY 242 AP-----VDNKNIAEVNLPESLETI-----SDYAFHLAKQI-----D 275
Db 1346 QIENOEKSLDKLAKSLDNKEKILVLDLKEKESLETRIKELNDIAYSNSKEMQTKNEN 1405
QY 276 LPDNLKAI-GELAFDQITQKLSLPQLMRLAERAFKSNHIKTIETFRGNSLKVIGEASF 334
Db 1406 LETKLSKTEKDLSTLNSKFTNETKILKDLI-----SDHEVSI-----SSLKVDLDKVK 1453
```


DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Putative transglycosylase.
 GN Name=yomi;
 OS Bacteriophage SPB2.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=66797;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Maue C.,
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF020713; AAC13005.1; -.
 DR F01, T12796; T12796.
 DR HSP, P00171; ILSP.
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002886; Peptidase_M23B.
 DR InterPro; IPR008258; SLT.
 DR InterPro; IPR010090; Tape_mae_Tp901.
 DR InterPro; IPR000189; Transglyc_AS.
 DR Pfam; PF01551; Peptidase_M23; 1.
 DR Pfam; PF01464; SLT; 1.
 DR TIGRFam; TIGR01760; tape_mae_Tp901; 1.
 DR PROSITE; PS00222; TRANSGLYCOSYLASE; 1.
 SQ SEQUENCE 2285 AA; 252310 MW; FF602C227754B357 CRC64;

Query Match 3.9%; Score 199.5; DB 2; Length 2285;
 Best Local Similarity 20.9%; Pred. No. 4.4;
 Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;

QY 81 EKKIKNLAKPREQELKAVNTES-EXQITSGSQLEQSKSLSLN----- 125
 DB 771 KKAKDDFQS-QQTNVEAITTKDSTDKLIQYKELQKVKESRSITDEEQEYLQVTOOL 829
 QY 126 -KTPVSTSNWEICDFITKNTLVGLSKSGVVKLSOT-DHLVLPQAA-----DQTQ 174
 DB 830 AQTFPAL-----VKGYSQNALTKNKELEKAIENKYLAKQETRDSAKKTFEDASK 885
 QY 175 LIQVASFAPTPDKTAIAEYTSRAGENGHISQLDVDGKEIINEGVFNFSYLLKKVTIPTG 234
 DB 886 EIKKSKDELQYKQ--IADYNDKGRPKWDLIADDDYKVAADKAK-QSMLKAQSDIESG 941
 QY 235 YKHIGQDAFVDNKNTAEVNLPSLET-ISDYAPAHALAKQIDLPNLKAIGELAFPDNQI 293
 DB 942 NAKVKDSVLSIANAYSSIDISNTLKTSTSD-----VNVKMLKDDLP-EELEKFSSSL 994
 QY 294 TGKLSLPQLMLAEAPKSNHIKTEPRGNSLKVIGEASFQDNDLSQMLPDGLEKTES 353
 DB 995 -QKLEKMQ-----KALDSGDEKAFD---NAK-----DLQSL-----LKEA 1024
 QY 354 EAFTGNPGDDHYNRNVLTWTKSGKNPSGLATENTYVNDPKSLWQSPSIDYTKWLEEDFT 413
 DB 1025 -----ETYSKSDS-----IDVFKMSFD--K 1043
 QY 414 YQKNSVTFPSNGKQKVRKNKLEIPKQHNGVTITEIGNAPRVNDPQNTLRKYDLSEV 473
 DB 1044 AQNKIKDG--DKSLSSVKSEVG-----DLGETLAEAGNEA---EDFGKK-----LKEA 1086
 QY 474 KLPSTIRKIGAPAFQSNLKSPEASDDLEIEKEGAPMNRITLKLKLVITIGDAAPHI 533
 DB 1087 LDANSVDDIKAAIKEMSDMQP---DSVDQVLNGDIPNN-----TKQVAPLND----- 1132
 QY 534 NHYIAVLPEVQIEGRSAFRQNGANNLIFPMGSKVTKTGEMAFLSNRLHLDLSQKQLT 593
 DB 1133 -----LLEKMAE-GKS-ISANEANTLI---QKQKELAQAIISIENGVVKINRDEVIQR 1180
 QY 594 ELPVQAFSD-----NALK-----EVLLPASKLTIRE----- 620
 DB 1181 KYKLDAYNDMVTYSNKLAKMTEVNNNAIKTLNADTLRIDSLKKLRKELKMDSEALSDELV 1240
 QY 621 -----AFKCNHLKQLE-----VASAL-SHTAFNALDDNDGDEQPD 654

DB 1241 KSINNVADAKKELKLEKMLQPGYNSQIEAMQSVKSALESYI--SASEEATSTQEMN 1298
 QY 655 NKVVVKTHTNSYALADGEHFIVDPDKLSSITVD-----LEKILKLEGL-----DYSTLR 704
 DB 1299 KQALVEAGTSLNWTDOQEKANEBTKTSMYVVDKYKEALEKNAEIDKYNKQVNDYPKYS 1358
 QY 705 QTTQTQFRD-MTTAGKALLSKSNLRQSEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKK 763
 DB 1359 Q-----KYRDAIKKEIKALQKKLQ--EAKLLAD-QIKSGNI-----TOTGIIVTST 1404
 QY 764 ATKNG--QLLERSINKAVLAVNNSAIKKANVKRLEKELDLTLGLVEGK-GPLAQTATWQ 820
 DB 1405 TSSGGTSSSTGSGYSGKYSYINSAASKYNV-----DPALIAAVIQESGNAKARGVG 1459
 QY 821 VYLKTLPLPLPYIYIGLVYFDKSGKLIYALDMSDTTIGEGQK-DAYGNPILNVDEDNEG 879
 DB 1460 AMGLMQLMPATAKSLGVNNAYDP-----YQVWVGTKYLAQLEKFGG--NVEKALAA 1511
 QY 880 HALAVATLADYEGLD-----IKTIL-----NSKLSQLTSTIRQVPTTAAHYHRAGIFQAIQN 928
 DB 1512 NA-GPGNVIKYGGIPPFKQETQYVYKIMANYSKLSATS-----SIASY----- 1556
 QY 929 AAEAEQLLPKPGTHSEKSSSESANGKDRGLQSNPKTNRGRHSAILPRTGSKG--SFVY 986
 DB 1557 -----TNSAFRVSSKYQGESGLRSP--HKGTDPAAKAGTAIKSLQSGKV 1601
 QY 987 GILGYTSTA 995
 DB 1602 QIAGYSKTA 1610

RESULT 84
 O31976
 ID O31976 PRELIMINARY; PRT; 2285 AA.
 AC O31976;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Yoni protein.
 GN Name=yomi; OrderedLocusNames=BSU21350;
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.-Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M.,
 RA Klein L.-M., Joris B., Karamata D., Kaeshara Y., Klaerr-Blanchard M.,
 RA Kleins C., Kobayashi Y., Koetter P., Koningstein G., Krogh S.,
 RA Kunano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,
 RA Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C.,
 RA Medigue C., Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S.,
 RA Nock M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,
 RA Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
 RA Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G.,
 RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
 RA Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
 RA Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Seror P.,
 RA Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,
 RA Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P.,
 RA Tognoni A., Tosato V., Uchiyama S., Vandenberg M., Vannier F.,

RA Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H.,
RA Weitzengger T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.P., Zumstein E.,
RA Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.",
RL Nature 390:249-256(1997).
DR EMBL; Z99115; CAB14053.1; -.
DR HSSP; P00717; ILSP.
DR GO; GO:004222; F-metalloendopeptidase activity; IEA.
DR GO; GO:006508; P-proteolysis and peptidolysis; IEA.
DR InterPro; IPR002886; Peptidase_M23B.
DR InterPro; IPR008258; SLT.
DR InterPro; IPR010090; Tape meas_TP901.
DR InterPro; IPR001189; Transglycyl_AS.
DR Pfam; PF01551; Peptidase_M23; I.
DR Pfam; PF01464; SLT; 1.
DR TIGRFAMs; TIGR01760; tape meas_TP901; 1.
DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
KW Complete proteome.
SQ SEQUENCE 2285 AA; 252310 MW; PF602C22754B357 CRC64;

Query Match 3.9%; Score 199.5; DB 2; Length 2285;
Best Local Similarity 20.9%; Pred. No. 4.4;
Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;

QY 81 EKIKNLAKPREOELKAVNTES-EKOITSGSQLEQSKESLSL----- 125
DB 771 KKAKDDFES-QQTVEAITTKDSTDKLIQQKELQKVKESRSLTSDEQEYLQVTOQL 829
QY 126 -KTVFSTSNWEICDFITKNTLVLGSLGSGVEKLSQT-DHLVLPQAA-----DGTQ 174
DB 830 AQTFPAL-----VKGYDSQGNAILKTNKELEKALIENTKEYLALKQETRSACKTFEDASK 885
QY 175 LIQVASFATPKKTAIAYTSRAGENGEISQLDVGKEIINEGEVFNLYLKKVTIPG 234
DB 886 EIKSKDELQYKQ--IADYDKGRPKWDLIADDDYKVAADKAK--QSMRLKAQSDIESG 941
QY 235 YKHGQDAFVDNKNTAEVNLPSLET-ISDYAFALHALAKQIDLPNLKAIGELAFDNOI 293
DB 942 NAKVKDSVLSIANAVSSIDISNLTSTISD-----VVKMLKDDLP-ELEKFPSSSL 994
QY 294 TKGLSLPQMLRAERAFKSNHKTIEFRGSLKLVIGEASFQDNLSQLMLPDGLEKIBS 353
DB 995 -GKLQEKMQ-----KALDSGDEKAFD--NAKK-----DLQSL----- 1024
QY 354 EAFNGPDDHNNRVVLTKSGKNPGLATENTYVNPDKSLWQSPEDITYTKWLEEDPT 413
DB 1025 -----ETYSKSDS-----IDVFKMSFD--K 1043
QY 414 YKNSVTGFSNGLQVKRKNLEIPKQHNGVTITEIGDNAPRVNDFQNTLRKYDLERV 473
DB 1044 AQNKIKDG--DKSLSVKSEVG-----DLGETLAEAGNEA--EDFGKK-----LKEA 1086
QY 474 KLFSTIRKIGAFAPSNLKSFEASDDEIEKEGAPMNNRIETLEKDKLVITIGDAAPHI 533
DB 1087 LDANSVDDIKAAIKEMSDAQF--DSVDVLNGDIFNN-----TKQVAPLND----- 1132
QY 534 NHIYAIVLPESVOEIGRSFRONGANNLIFMGSKVTLGEMAPLNRLHLDLSQKQIT 593
DB 1133 -----LLEKMAE-GKS-ISANEANTLI---QKQELAQAIISIENGVVKINRDEVIKOR 1180
QY 594 EIPVQAFSD-----NALK-----EVLLPASIKTIRE----- 620
DB 1181 KVKLDAYNDVMTYSNKLMTENVNAIKTLNADTLRIDSLLKLRKGRKLDMSAEISDLEV 1240
QY 621 -----AFKKNHLKQLE-----VASAL-SHIAFNALDNDGDGEQFD 654
DB 1241 KSNINNVADAKKELKEKWLQPGGYSNSQIEAMQSVKSALSYI--SASEATSTQENY 1298
QY 655 NKVVKVTHNSVALADGEHFI VDPDKLSSTIVD-----LEKLIKUEGL-----DYSLTR 704
DB 1299 KQALVEAGTSLENWTDOQEKANEETKTSMYVVDYKVALEKYNABEDKYNQVNDYPKYS 1358

QY 705 OTTOTOFRD-MTTAGKALLSKSNLRQGBKQKFLQBAQFPLGRVDLDKAIKAEKALVTKK 763
DB 1359 Q-----KYRDAIKKEIKALQOKKLLMQ-EQAKLLKD-QIKSGNI-----TQYGI VTST 1404
QY 764 ATKNG-QLLERSINKAVLAYNNSAIKXANVRLEKELDLLTGLVEGK-GPLAQATWVOG 820
DB 1405 TSSGGTSPSTGSGYSKYSSYNSAASKYV-----DPALIAAIVQESGFNAKARSGVG 1459
QY 821 VYLLKTPPLPEYIYGLNVPYFDKGLIYALDMSDTIGEGOK-DAYGNPILNVDENECY 879
DB 1460 AMGLMQLMPATAKSLGVNNAYDP-----YQNVMGGTKYLAQQLKFGG--NVEKALAA 1511
QY 880 HALAVATLADYEGLD-----IKTIL-----NSKLSQLTISRQVPTAAVHRAGIFOAIQN 928
DB 1512 NA-GPGNVIKYGGIPFPKETQNVYKIMANYSKSLSSATS-----SIASY----- 1556
QY 929 AAEAEQLLPKPQTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG--SFVY 986
DB 1557 -----TNSAFRVSSKYQOESGLRSP--HKGTDFAAKAGTAIKSLQSGKV 1601
QY 987 GILGYTSA 995
DB 1602 QIAGYSKTA 1610

RESULT 85

QYWE8
ID Q7YWE8 PRELIMINARY; PRT; 3256 AA.
AC Q7YWE8;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Normocyte binding protein 2b.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI TaxID=5833;
RN (1) SEQUENCE FROM N.A.
RC STRAIN=7G8;
RA Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,
RA Corredor V.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY138501; AAN39448.1; -.
SQ SEQUENCE 3256 AA; 383108 MW; 479FF52584BD8194 CRC64;

Query Match 3.9%; Score 199.5; DB 2; Length 3256;
Best Local Similarity 19.2%; Pred. No. 7.3;
Matches 209; Conservative 177; Mismatches 356; Indels 347; Gaps 52;

QY 32 ILKQTOASSISGADYAESGSKLKINETS-----GPVDDTVTDLFSDKRTTPEKIDN 86
DB 2026 LVKDCKELELSTALY-----DLKIQTTSVINRENDISNNI-DIVSNKLEINAIQYN 2077
QY 87 LAKGPREQELKAVTENTESEKQI--TSGSQLEQSKES-----SLSLNKTVPSTSNWEICDFTT 141
DB 2078 FEK-----YKEIFDNVEYKTLDDTKNAYIVKAAELKQVDINTKE-----DLDI 2123
QY 142 KNTLVGLSGKGEKLSQTDHLVLPQAADGTQLIQVASFAPTPDKKTAIABYTSRAGEN 201
DB 2124 YPNLDLELEKS-----LTLSSNEMETKTIVQ-----NSYNSFSDIN 2159
QY 202 GEISQLDVDGK-----EINEGEVFN-----SYLLKKVTIPTG----- 234
DB 2160 KNINDIKEMKTLIPMLDELLNEGHNDISLYNFIIRNIQIKIGNDKINIREQENDTNC 2219
QY 235 -----YKHIGQDAFVNK-----NIAEVLNPSLETISDYAFALHALKQI 274
DB 2220 PEYIQNNYFNFKSDISIFNKYDDHDKVDNYISNNIDVKNKNSL--LSEH-----VINAT 2272
QY 275 DLPDN-LKAIGELAFDNOITGKLSLPR--QLMPLAERAFK-----SNHKTIEFR--- 322
DB 2273 NIIENIMTSIVEI-----NEDTEMNLEETQDKLELYENFKKEKNIINNRYKIVHFNK 2328


```

Db 665 RTLSGFAENSLSEIEPEKALEPLTSLNFDLSQNNLALLPKTIEKLRALSTIVASGRNHT 724
Qy 583 HLDLSEQKOLTRI PQAFPSDNALKVELLPASLKTIREBAFKNH-LKQLE-----VASALS 637
Db 725 RLDNISFRLPKSLVDLSTNEISN-LPNGI-----FKONQOLTKLDFNNLLTQVE 775
Qy 638 HIAFNALDNDGDEQFD-----NKVVVKT-----HNSYA 667
Db 776 EGVFPDVTNLNDFVFNQIKSVSPKVRALIGQHLTPQKHIAKLASLDGKIKYHQAFS 835
Qy 668 LADGEHFIVDPKLSSTI-----VDLEKILKLE-----GLDYSTLRTQTQFRDMTTAGKAL 721
Db 836 LLDLYW---EQTNSAIDKELVSEYQQLQEKSGDTSVLLNDMQVMSIVIOLQKKA 892
Qy 722 LSKSNLRQEKQKFLQEAQFFLGRVDL-DKATAKAEKALVTKKATKNGQLLERSINKAVL 780
Db 893 SNGQVTVDEKLLSNDPKDLDLTFBESLKDPPGYRIRKALITTKKFAATQKEHIYLTSDILV 952
Qy 781 AYN-----NSAIKKANVRLEK-----ELDLTLGLVEGKGPLAAQATWV--Q 819
Db 953 AKGPHSHQKDLVENGRLNQQLRDGIYVLNASMLKTDLASESNSKAINHRVTLVVK 1012
Qy 820 GYVLKTLPLPBEY-----YGLNVYF-----DKSKLI-----YALDMSDT 856
Db 1013 GVSYLEV-----EPRGIKVGKMLGYLGELSYFVDGYQYRDLQAGKVPVRTKAEVWSYFTDV 1067
Qy 857 IGEQKQDAYGN-----PLNVDEDNEG-----YHALAVAT 886
Db 1068 TGLPLADRYGKVPKVLKMLKLEQAKDGLVPLQVFPIMDAISKSGSLQTVFMKLWAS 1127
Qy 887 LADYEGLODKITLIN-SKLSQLSITSIRQV-----PTAAYHRAGIFQAIQNA 930
Db 1128 LTTAKAVKVTNNQENSHLSTDLQKGPQRQOEKTPSPSAAATGIANLTDLLAKA 1187
Qy 931 A-----EABQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNKGRHSAILP 976
Db 1188 TQCSQETSQTDITDKAEKQLQVRDH---QTSIEGKTAKDTKTKSKDKHRSNQQSNGE 1244
Qy 977 RTGSKGSFVYGLGYTSVAL 996
Db 1245 ESSRYHLIAGLSSPMIVAL 1264

```

RESULT 87

```

Q708A9
ID Q708A9 PRELIMINARY; PRT; 1842 AA.
AC Q708A9;
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE AGCP15288 (fragment);
GN Name=agCG50148; ORFNames=ENSANGG00000008609;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAA801008944; EAA10144.1; -.
FT NON TER 1842
SQ SEQUENCE 1842 AA; 208361 MW; 978A013DB2ADD5FC CRC64;

Query Match 3.9%; Score 199; DB 2; Length 1842;
Best Local Similarity 19.3%; Pred. No. 3.5;
Matches 226; Conservative 180; Mismatches 366; Indels 402; Gaps 55;

```

```

Qy 2 KHLKTKVALTLTTSVVVTH-----NOBFVSLVKE-----PILKQTOA 38
Db 304 ERSITTEITTEIVDHTDLRELQENESLRLRELDTVKARAERAERDKSILLRRLAS 363
Qy 39 SSSISGADYASSGSKSLKINETSQVDDTVTDLPSDKRTPTEKIK-----DNLAGKPR-- 92
Db 364 IDTVSNKTAASEALKOOKVNEQKOLID-----DLQDEKKFLTKLSELSADMSARGGRI 419
Qy 93 -----EOLKAVTENTSEB-KOITSGSOLEQSKESLSLNTKV-P 129
Db 420 BEQLRQKLEQAEATCEBELMDENEELKRELNMWETIEEWHNDFREEQADEVASLKKELDQ 479
Qy 130 STSNWEICDFITKGNLTGLSKSGVEKLSQTDHLVLPQAAADGTOLIQVASFAFTPKKT 189
Db 480 TTKNCRILSF-----KLKSDR-----RIEQLS-----EKA 506
Qy 190 AIAEYTSRAGENGE-----ISQLDVDGKEIINE-GEVFNISYLLKKVTITPT 237
Db 507 AL-----GASGDIAAKIKQLE-DELKVSNEVARRLQAELESSTAPTSTPTKTFPSLGN 558
Qy 238 IGQDAFVDNK-----NIAEV--NLPSLETISDYAPFAHLAKQIDLDPNLKA 282
Db 559 IGKTSADSKMSRASLFRGGSQEDPAQLRDLQDSLE-----READLREQUKY 606
Qy 283 IGEAFFDNOITGKL-----SLPRQLMRLAERAFKSNHI-----KTIEFRGNSLK 327
Db 607 ABEBAENLRKRSRVEDENDSLVMQLKKWATKATKTSKSI FSKARKLSPATSRSNADPAP 666
Qy 328 VIGBASPOD-----NDLSQMLPDGLEKIESEAPTGNPGDDHYNRVV----- 370
Db 667 VEKDEGISDERDPAELRLQLEINEQEMAVLRRKMELEHE-----NKHAREQIKDQLE 719
Qy 371 -LWTKS---CKNPSGLATENTYNP--DKSLWQESPEID--YTKWLEEDFTYQKNSVTG 421
Db 720 NLAFTKTELGRVPSALGSKGVDPLEBKITVMEDEISELRKKLIEKREFERLQ--- 776
Qy 422 FSNKGLQVKKNKNLEIPKQHNGVTITIGD--NAFRNVDFQNKTLR-----KY 468
Db 777 -AEMALAKSGKSLSKTSKSLDALTEQOQDLKQLOVIEQEAIVRAKTQSLQEDNEKY 835
Qy 469 DLEEVKLPSTTRKTGAPAFQSNNLKSPASDDLEBI-KEGAFMNNRIE-----TLEKLD 521
Db 836 QAEIKKLQTKAKPAGA---TSAADTKLTDSEIEQLEKEKEQLEGLKRIVQESTSOLPM 891
Qy 522 KLV-----TIGDAAFHINHIYAILVPESVQETIGRSAFRQNGANNLIFMGSK 567
Db 892 RLAKSPNDMHTYQLQRKMWEDSELEIDELRALV-----GKA-----GAMNIATLEKE 938
Qy 568 VKTLGEMAFLSNRLEHLDLSQKQLTEIPVQAFSDNALKEVLLPASLKTIREEAPKQNH 627
Db 939 KK-----RLB-AELAEKQOQE-----KLITTELSTLKKETLEQOTS 973
Qy 628 KOLEVASALSIIAFNALDNDGDSQFDNKKV-----VKTHNSVAL----- 668
Db 974 KLEBAQTVERLEDENRKQDKIKTLEDKTRVNTTMTKTBASSKSLLEIQLKAEKEHTG 1033
Qy 669 --AGEHFIVDPDKLSSTIVLEKILK-----IEGLDY 700
Db 1034 TERDLEKVRKEKTKLDNRIISOLEKELQLSKNAELIKESLREMAALKSKSATGAEDSES 1093
Qy 701 STLRTQTTQTQPRDMTT-----AGKALLSKSNLRQGEKQKFLQEAQFFLGR--VDLDK 750
Db 1094 SKILLQDLKKQNEELTTEIHQOSRKFEELLOKHETM--EEHLVTVQAOALAAREKLOELD 1151
Qy 751 ---AIAKAEKALVTGKATKNGQLLERSI---NKAVALYNNNSAIKKANVYLRLEKLDLITG 804
Db 1152 LKNKLLQAE-AIETRLVKENTNMSRRLVEVQKLTAAETNDSRYASIE-LEKN----- 1203
Qy 805 LVEGKGPLAAQATMVQGVYLLKTPLPYIYGLNVYFDKSGKLIYALDMSDTIGEGQKDA 864
Db 1204 -----RMKTALE-----DKQH-----DYEKLAKENEMNA 1227
Qy 865 YGNPILNVDEBDEGYHALAVATLADYEGLDIKITILNSKLSQLSITRQVPTAAYHRAGIFQ 924

```

```
Db 1228 Y--QVNQLKKNEDLR-----KLLDYERIN-----KAQRTLSERNAHLEQ 1266
Qy 925 AIONAAABEQLLPKPGTHSKSSSSSSANSKOR 958
Db 1267 ELKSVYTKLEA-----SEMWKSEVAATRLR 1292

RESULT 88
Q7WS91 PRELIMINARY; PRT; 2186 AA.
AC Q7WS91;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE FmbB-like protein.
GN Name=mip-homologue;
OS Staphylococcus aureus;
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1180;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=PC3, Michigan, New Jersey, St Luke, PC1, Slovenia6096, and
RC SMH2;
RA Wootton M., Avison M.B., Bennett P.M., Howe R.A., MacGowan A.P.,
  Walsh T.R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ567420; CAD98830.1; -
DR EMBL; AJ567421; CAD98831.1; -
DR EMBL; AJ567422; CAD98832.1; -
DR EMBL; AJ567423; CAD98833.1; -
DR EMBL; AJ567424; CAD98834.1; -
DR EMBL; AJ567425; CAD98835.1; -
DR EMBL; AJ567426; CAD98836.1; -
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR011591; Botulinum.
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR011490; FIVAR.
DR InterPro; IPR005877; Gpos_YsIRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF07564; DUF1542; 17.
DR Pfam; PF07554; FIVAR; 1.
DR Pfam; PF04650; YsIRK_signal; 1.
DR ProDom; PD001963; Botulinum; 1.
DR TIGRfams; TIGR01167; LPXTG_anchor; 1.
DR TIGRfams; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 2186 AA; 238445 MW; 0CC84836F1868A4F CRC64;

Query Match 3.9%; Score 199; DB 2; Length 2186;
Best Local Similarity 19.2%; Pred. No. 4.4;
Matches 217; Conservative 182; Mismatches 387; Indels 342; Gaps 58;

Qy 20 HNOEVSLVKEPILKQTOASSISGADYAESGSKLKNETSGPVDVTDLPSD---- 75
Db 1208 YNAKLAENATPDATNDEKNAINTLNQDROAIESIKQANTNAEVDAQVAENNDIV 1267

Qy 76 -----KTTPEKIDNIAKPREQELKAVTENTSEKQITSGSQLEQSK-----SLSLN 125
Db 1268 QVDVVKKQAARDKITAFAK--RIEAVKQTPNATDEEQ--AAVQINQLKQAINQIN 1324

Qy 126 KT--VPSTSNWEI-----CDFITKNTLVGLSKSGVEKLSQTDHLVLPQAADGTQL 175
Db 1325 QTNQDVTITNQAVNAIDNEAEVVVVKPKAIDIEKAVKEKQQQIDNSL-----DSTDNEK- 1380

Qy 176 IQVASFAFTPKTAIAEYTSRAGNGRISQLDVDGKEIINEGEVFNYSLLKKVT--Ipt 233
Db 1381 -EVASQALAKEKALAA-IOAQTNQVQNAATNGVSAIK-----IIQPTKVKKPA 1430

Qy 234 GYKHIGQDAFVNDKNIAENVLPE-----SLETISDYAFALH-----KQID 275
Db 1431 AREKINOKA---NELRAKINQDKATEERQVALDKINE--FVQAMTDTITNRTNQVD 1485
```

```
Qy 276 LPDNLKAIGELAFDNOITCKLSLPRLMLAER-----APKSNHIKTIERGNLSKV 328
Db 1486 -DTTSQALDSIA-----LVAPEHIVRAAADAVKQYEAQKQIEQAEHATDEKQ 1535

Qy 329 IGBASFODND-----LSQLMLPDGLEKIESEAFTHGDDHYNNRVLVLTWTKSGKNSGL 382
Db 1536 VALNQLANNEKLALONINQAVTNDVKRVETNGIATLKG---VQPHIVI-----KPAQQ 1587

Qy 383 ATENTYVNPDKSLWQESP-----EIDYTKWL-----BEDFTYQKNSVTFGSN--- 424
Db 1588 AIRATAENQVESI-KOTPHATVDELDEANQLISDTLKQAQOEIENTNQDAADVTDVRQTI 1646

Qy 425 KGLQKVK-----RNKN-----LEIPKOHNGVTIETGD--NAPRNV 459
Db 1647 KAIEQIKPVRRKRAALDSIENNNKQLDAIRNTLDTTODERDVAIDTLNKIVNTIKNDI 1706

Qy 460 FOKTLRKYDLE-----VKL-----PSTIRKIGAFAFOSNNIKSPASD-DLEEI 504
Db 1707 AQKNTNAEVRTETDGDNDNIKVLPKVQVPAARQSVGVKAEQAQNAL--IDQSDLSTEE 1764

Qy 505 KEGA-----PMNRIETLEKDKLVITIGDAAPHINHIYAVLP-----ESVQEIG 549
Db 1765 RLAAKHLVQALNQAIQDINHADKTAQVNDQSIDAQNIISKIKPATTVKATALQIQNIA 1824

Qy 550 RSAPFRQGAN-----LIPMSGKVK---TLGEMAF-----SNRLEHLD 585
Db 1825 TNKINLIKANNEATDEEQNTAIAQVEKELIKAKQIATASAVTNADVAYLLHDEKNEIREIE 1884

Qy 586 -----LSEOKOLTEIPVQAFSDNALKEVLLPASLKTIREEPKKNHLKOLEVASALSHI 639
Db 1885 PVNRKASAREQLTTL---FND---KQAIEANIQATVEE--RNSILAQLQ---NIYDT 1932

Qy 640 AFNALDDNDGDEQPDN--KVVVTHHNSYALADGEHFIVDPDKLSSTIVDLKILKLEIG 697
Db 1933 AIGQIDQDRSNAQVDKTAASLNQTIHDL-----DVHPFKKPDAAKTINDDLARVTALVQ- 1986

Qy 698 LDYSTLRQTTQTQFRDMTTAGKALLSKNSLRQSEKQKFLQEAQPFGLGRVLDLDAKIAAEK 757
Db 1987 -----NYRKVSNRNK-----ADALK 2001

Qy 758 ALVTKATKNGQLLERSIN--KAVLAYNNSAIKKNVKELEKELDLTLGLVEGKGLAQ 814
Db 2002 AITALKLOMBELKTARTNADVDVAVLKRFNVALSDIEAVITEKENSLLR--IDN---IAQ 2056

Qy 815 ATMVQGVYLLKTPLPPEYIYGLNVYPDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDE 874
Db 2057 QT-----YAKFKAIATPEQLAKVKVLIDQ-----YVAD-----GNRM--IDE 2091

Qy 875 DNEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVYHRAGIFQAIQNAAEAE 934
Db 2092 D-----ATLN-----DIKQHTQFIVDEILAIAK-LPAEA-----TKVSPKEI 2126

Qy 935 QLLPKPGTHSEKSSSSSSANSKORGLQSNPKTNRGRHSAILPRTGSKG 982
Db 2127 QPAPKVCTPIKGETHESRKVEK-----LPNTGSEG 2158

RESULT 89
Q99TB0 PRELIMINARY; PRT; 2186 AA.
AC Q99TB0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE SAI577 protein.
GN OrderedLocusNames=SAI577;
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=N315;
RC MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
EX
```


RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguni A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaiko C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayaashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant *Staphylococcus*
RT aureus".
RL Lancet 357:1225-1240(2001).
DR EMBL: AP001135; BAB42845.1; --
DR F1R; H89960; H89960.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR011591; Botulinum.
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR011490; FIVAR.
DR InterPro; IPR005877; Gpos_Y8IRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF07564; DUF1542; 17.
DR Pfam; PF07564; FIVAR; 1.
DR ProDom; PD001963; Botulinum; 1.
DR TIGRfam; TIGR01167; LPXG_anchor; 1.
DR TIGRfam; TIGR01168; Y8IRK_signal; 1.
DR Complete proteome.
SQ SEQUENCE 2186 AA; 238448 MW; OCC84836F1868A4F CRC64;

Query Match 3.9%; Score 199; DB 2; Length 2186;
Best Local Similarity 19.2%; Pred. No. 4.4;
Matches 217; Conservative 182; Mismatches 387; Indels 342; Gaps 58;

QY 20 HNDVFLVKEIPKQTOASSISGADYAESGSKLKNETSGPVDVDTLDFSD--- 75
DB 1208 YNKLAEINATPDATNDEKNAINTLNQDRQAISIKQANTNAEVDQAATVAENIDAV 1267

QY 76 -----KRTPEKIDNKLAKPREQELKAVTENTESEKQITSGSQLEQSK-----SLSLN 125
DB 1268 QVDVVKQAARDKITAEVAK--RIEAVKQTPNATDEEK--AAVQINQLKQQAQINQON 1324

QY 126 KT---VPSTSNWEI-----CDFYKGNLTGLSKSGVEKLSQTDHLVLPQAADGTQL 175
DB 1325 QTNQDQVDTTQNAVNAIDNVEAEVVIKPAIAIDIEKAVEKQKQIDNSL---DSTDNEK- 1380

QY 176 IQVASFAPDPDKTAAETYSRAGENGESQLDVGKEIINEGEVFNLSLLAKVTV--IPT 233
DB 1381 -EVAQALKEKEKALAA-IDQAQNSQVNAATNGVSAIK-----IIQPTKVRPA 1430

QY 234 GYKHIGQDAFVDNKNIAEYNLPE-----SLETISDYAFALAL-----KQID 275
DB 1431 AREKINQKA--NELRAKINQKAEATAEERQVALDKINE--FVQAMTDTITNRTNQVD 1485

QY 276 LPDNLKATGELAFNQITGKLSLPQLMRLAER-----AFKSNHKTIFRGNLSKV 328
DB 1486 -DITTSQALDSIA-----LVAPEHIVRAAADAVKQYEAQKEIEQAESHATDEEQ 1535

QY 329 ICEASFQND-----LSQLMLPDGLEKLESEAFNCGDDHNNRVNVLTKSGKNPSGL 382
DB 1536 VALNQLANNEKALQINQAVTNNDVKRVETNGIATLKG---VQPHIVI-----KPEAQ 1587

QY 383 ATENTVYVNPDKSLWQESP-----EIDYTKWL-----BEDFTYQNSVTFGSN--- 424
DB 1588 AIKATAENQVESI-KDTPHATVDELDEANQLISDTLKQAQQEIENTNQDAATVDVNRQTI 1646

QY 425 KGLQKVK-----RNKN-----LEIPKQNGVITITEIGD--NAPRNV 459
DB 1647 KAEIQIKPVRKRAALDSIEENKNQDLAIRNTLDTQDERDAVIDTLNKIVNTIKNDI 1706

QY 460 FQNKTLRKVDLEE-----VKL-----PSTIRKIGAFQSNILKSFEASD-DLREI 504
DB 1707 AQNKTAENAEVDRTETDGNNDIKVILPKVQVQPAARQSVGVKAEQAQNAL--IDQSDLSSTEE 1764

QY 505 KEGA-----FMNRIETLELKDCLVTIGDAAFINHIVAILVP-----ESVQBIG 549

DB 1765 RLAAKHLVEQALNQALDOINHADKTAQVNOQSDIAQNIISKIKPATTVKATALQOIONTA 1824
QY 550 RSAPFQNGANN-----LIFMGSKVK---TLGMAFL-----SNRLEHLD 585
DB 1825 TNKINLIKANNEATDEQNIATIAQVEKELIKAKQOIASAVTNADVAYLLHDEKNEIREIS 1884
QY 586 -----LSEQKQTEIPVQAFSDNALKEVLLPASLKTIREEAFFKNHKLQLEVASALSHI 639
DB 1885 PVINKASAREQUTTL-----FND---KKQATEANTQATVEE--RNSILAQLQ---NIYDT 1932
QY 640 AFNALDDNDGDQFDN--KVVVKTHNSYALADGSHFIVDPDKLSTSTIVDLKILKILIEG 697
DB 1933 AIGQIDQDRSNAQVDKTASLNLQTHDL-----DVHPICKPDAEKTINDDLARVLTALVQ- 1986
QY 698 LDYSTLRQTTQTFQDMTGTAGKALLSKSNLRQGEKQKFLQEAQFPLGRVLDLDAKAKAEK 757
DB 1987 -----NYRKVSNNRK-----ADALK 2001

QY 758 ALVTKKATKNGQLLERSIN---KAVLYNNSAIKKNVKKLEKELDLTLGLVEGKGLAQ 814
DB 2002 AITALKQWDEBELKTARTNADVDVAVLKRPNVALSDIEAVITEKNSLLR--IDN---IAQ 2056

QY 815 ATMVGQVYLLKTPPLPEYIYGLNYPYFDKSGKLIYALDMSDTIGEGKDAYGNPILNVDE 874
DB 2057 QT-----YAKFKAIATPEQLAKVKVLIQ-----YVAD-----GNRM--IDE 2091

QY 875 DNEGYPHALAVATLADYEGLDIKTILNSKLSQTSIRQVPTAAYHRAGIFQATONAAAEAE 934
DB 2092 D-----ATLN-----DIKQHTQTFIVDEILAIAIK-LPAEA-----TKVSPKEI 2126

QY 935 OLLPKPGTHSKSSSESANGKDRGLQSNPKTNRGRHSAILPRTGSKG 982
DB 2127 QPAPKVCPTPIKEETHESRKEVE-----LPNTGSEG 2158

RESULT 90
Q6GFU5
ID Q6GFU5 PRELIMINARY; PRT; 2189 AA.
AC Q6GFU5;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Putative surface anchored protein.
GN Name=sac; OrderedLocusNames=SAR1841;
OS *Staphylococcus aureus* (strain MRSA252).
OC Bacteria; Firmicutes; Bacillales; *Staphylococcus*.
RX NCBI_TaxID=282458;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-15213324; DOI=10.1073/pnas.0402521101;
RA Holden M.T.G., Fell E.J., Lindsay J.A., Peacock S.J., Day N.P.J.,
RA Enright M.C., Foster T.J., Moore C.E., Hurst L., Atkin R., Barron A.,
RA Bason N., Bentley S.D., Chillingworth C., Chillingworth T.,
RA Churcher C., Clark L., Cronin C., Cronin A., Doggett J., Dowd L.,
RA Feltwell T., Hance Z., Harris B., Hauser H., Holroyd S., Jagels K.,
RA James K.D., Lennard N., Line A., Mayes R., Moule S., Mungall K.,
RA Ormond D., Quail M.A., Rabinowitch E., Rutherford K.W., Sanders M.,
RA Sharp S., Simmonds M., Stevens K., Whitehead S., Barrall B.G.,
RA Spratt B.G., Parkhill J.;
RT "Complete genomes of two clinical *Staphylococcus aureus* strains:
RT evidence for the rapid evolution of virulence and drug resistance.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791(2004).
DR EMBL; EX571856; CAG40832.1; --
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR011490; FIVAR.
DR InterPro; IPR005877; Gpos_Y8IRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF07564; DUF1542; 17.
DR Pfam; PF07554; FIVAR; 1.
DR Pfam; PF04650; Y8IRK_signal; 1.

Db 1900 DMEEEKVVKIKQIITE-----ESEEELEISKDPT--TSHSEKPNIEBQSVN 1943

Qy 526 IGDAAPHIN-----HIY-----AIVLPESVQIEGRSAFRONGANNILFMGSKVKTIG 572

Db 1944 IVDKNEINVOIEKNVQNEQEGDPIILLEEQKNIAILENQKYNPNQSLSHKERTILL 2003

Qy 573 ENAFSLNRLE-----HLDLSEQKQLTEIPVQAFSDNALKEVLLPASLTKIIEEAPKQKH 626

Db 2004 EYDDEGSDMTDCLTSELNKNKCDISIIPASNTDNKLN-----KDITE---NKDD 2052

Qy 627 LKQLEVASALSHTAFNALDDNGEQFNKVVVVKTHNS-----YALADGE 672

Db 2053 FSEIEKSGVEIHENGKOLLNKAESDDVPVQNKIEHDSENAGVIDQYRYRTDYBENKV 2112

Qy 673 HFIVDPDK-----LSTIVDLKELKILIEGLDYSTILRQTQTQFRDMTTAGKA 720

Db 2113 NFVGLPGKPEERSDVTGRSGVPSVDAPFK--KLHEPLD-----2149

Qy 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKAT-----KNGOLLERSIN 776

Db 2150 ----SNEKNKEELNVSKETIEYPVGLDILGIDSESSDSVTIYKKPTDTFVENVHVLSDIK 2205

Qy 777 KAVLAYNNSAIKANKVRLKELDLTLGLVE--GKGPLAQATWQGVYLLKTPLPPEY 834

Db 2206 S----LFENEKTVGLNEQKETEKNIAAGMIDVLNKNIDDDQSVR---LSTSL---YH 2254

Qy 835 IGLNVYFDKSGKL---IYALDMSDTTIGEGQKAYGNPILNVD-----EDNEGVA----- 881

Db 2255 KGLD--DOOSGRSLTALYDKGLDDOYHLHDSYNNINNIKNKVOCKENCGNGVSSF 2312

Qy 882 LAVATLADYE-----GLDIKTILNSKLSQUTSIRQVPTAAYHRAIGFQAIQNAAEQOL 936

Db 2313 VDTTLRDVKQKEHTEETKIENKIDNTEVDGVDITSDNRSS--FISVDNVDEKKET 2371

Qy 937 L-----PKPGTHSEKSSSESANSKDRGLQSNPK 965

Db 2372 IINEKNKDPKPSRKEEKKVKHDDSENDNNLSNIK 2407

RESULT 92

Q8NWQ6 PRELIMINARY; PRT; 9904 AA.

ID Q8NWQ6

AC Q8NWQ6; 01-OCT-2002 (TReMBLrel. 22, Created)

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)

DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)

DE Ebb protein.

GN Name=ebb; OrderedLocusNames=MW1324;

OS Staphylococcus aureus (strain MW2).

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI_TaxID=196620;

RN [1]_TaxID=196620;

RP SEQUENCE FROM N.A.

RC STRAIN=MW2;

RX MEDLINE=22040717; PubMed=12044378; DOI=10.1016/S0140-6736(02)08713-5;

RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,

RA Nagai Y., Iwano N., Asano K., Naimi T., Kuroda H., Cui L.,

RA Yamamoto K., Hiranatsu K.;

RT "Genome and virulence determinants of high virulence community-

RT acquired MRSA.";

RL Lancet 359:1819-1827(2002).

DR EMBL; AP004826; BAB95189.1; --

DR GO; GO:0016020; C.membrane; IEA.

DR InterPro; IPR011439; DUF1542.

DR InterPro; IPR010983; EF_Hand.

DR InterPro; IPR011490; FIVAR.

DR InterPro; IPR002988; GA.

DR InterPro; IPR005877; Gpos_YSRK.

DR InterPro; IPR006530; YD.

DR Pfam; PF07564; DUF1542; 7.

DR Pfam; PF07554; FIVAR; 53.

DR Pfam; PF01468; GA; 48.

DR Pfam; PF04650; YSRK_signal; 1.

DR TIGRFAMs; TIGR01643; YD repeat 2x; 1.

DR TIGRFAMs; TIGR01168; YSRK_signal; 1.

KW Complete proteome.

SQ SEQUENCE 9904 AA; 1068489 MW; CE96908F75EA56B2 CRC64;

Query Match 3.9%; Score 199; DB 2; Length 9904;

Best Local Similarity 19.1%; Pred. No. 36;

Matches 223; Conservative 175; Mismatches 451; Indels 316; Gaps 49;

Qy 1 MKKHLKTLVALTLTTVVVTHNOEVF-----SLVKEPILKQTQ---ASSISGADYA 48

Db 4502 LKEQVQS--ATTLEGVQTVKNSSQTLNTAMKGLRDSIANEATIKAGQNYTDASPTRNNEYD 4560

Qy 49 ESSGSKLKINETSQV--DDTVTDLFSDKRTTPEKIKDNLAKGPREQELKAVTENTESE 106

Db 4561 SAVTAAKALINQTSNPTMEPTIT-----QATSVTTKE 4594

Qy 107 KOITSGSQLEQSKESLSLNTVPSTSNWEICDPIYK---GNVLGLSKSGVE--KLSQT 160

Db 4595 HALANGAQNLAQAKTTAKNNLNLTSINNAQDALTRSIDGATTVAGVNOETAKATELNA 4654

Qy 161 DHLVLPQAADGTQLOVASFA-FTPDKKTAI-----AEYTSRACENGISQLD--- 208

Db 4655 MH-SLQNGINDETQTKQKYLDAEPNKKSAVDQAVNAAKAILTKASGQNVDKAAVEQAL 4713

Qy 209 -----VDGKEIINEGEVENSYLLKKVTIPTGYKHI--GQDAFVQNK-----NIAEV 252

Db 4714 QNVNSTKTALNGDAKLNKAAKAAKQTLGLT-----HINNAQRNALDNEITQATNVEG 4767

Qy 253 NLPESLETISDYFAHLKALKQIDLDPNL-----KAIGELAFFNQITG 295

Db 4768 NTVKAAQQLDCAMGOLETSIRDKDTTQLSQNYQDADAKRTAYSOAVNAATAILNKTAG 4827

Qy 296 ----KLSIPQLMRLAERAFKSNHIKTIEFRGNSLKVIGEASFQNDLS-----340

Db 4828 GNTPKADVERAMQAVQAANTALNGIONLE---RAKQAAANTAITNASDLNTKQEKALKQV 4884

Qy 341 ----OLMLPDGLEKTESEAFNGPDGHYNNRVLTGSKGNPSGLATENTVVPD----392

Db 4885 TSAGVSAANGVEHTATE-----INTAMTKRAIADKADTKTSGNVYNADANKR 4934

Qy 393 -----XSLWOESPEIDYTKWLEBDFYQKNSVTGFSNGKLOKVKRNKNLEIPKQH 442

Db 4935 QAYDEKVTAAESI VNGTPTTLT---PSDVT---NAAQVNTAKTQ-LNGNHNLEVAQN 4987

Qy 443 ----NGVT-----ITEIGDNAFRVDFQNKTLRKYDLEEVKL 475

Db 4988 ANTAIDGLTSLNGPOKAKLKEQVGOATTLPNVQTVRDNA-QTLNTAMKGLRDSIANEATI 5046

Qy 476 PS-----TIRK--ICAFAPQSNLKSPEASDLEIEKEGAF---MN 511

Db 5047 KAGQNYTDASPNRSEYDSAVTAAKAILGQTTSPMNAQETINQAKDQVTAQKQALNGQEN 5106

Qy 512 NRIETLELKLVTIGD-----AAPHINHIYAIVLPESVQIEGRSAFRQNG 557

Db 5107 LRTAQTNKAKOHLNGLSLDTNAKQAAKQIEGATHVNEV-----TQAQNN 5151

Qy 558 ANNLFMGSKVKT-LGEMAFLSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKT 616

Db 5152 ADALNTAMTNLKNIGIQDQNTIKQGVNFTDADEAKENAYTNVQAEQILNKAQGTAKDX 5211

Qy 617 IREBAFK-----KNHLKQLEVASALSHTAFNALD-----DNDGDGEQFNKVVVVKTHNSY 666

Db 5212 NVESALQNVQRAKNELNGQNVANAKTAKAKNNLNTSINNAQKALKSQI-----5262

Qy 667 ALADGEHFIVDPDKLSSTIVLEKIL-KLIEGLD-----YSTLRQTTQTQFRDMT 715

Db 5263 ---EGATTVAGVQVSTTASELNTAMSNLQGINDEAATKAAQKYTDADRDQKQAYNDV 5319

Qy 716 TAGKALLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKATKNGOLLERSI 775

Db 5320 TAAKTLDDKT-AGTNNENKAAVEQA---LQRVNTAKTALNGDARLNEAKNTAKQV---5370

Qy	575	AFLSNRLHDL	SEQKQLTEI	PVQAFSDNALKEV	----	LPASLTKTIREBAPKQHLK	628	
		:	:	:	:	:		
Db	1116	KENBEHLKEE	KIQLEKATET	KQOJLSNLRANLE	SELEKHEHDLAQLKCYEQI	ANKERQY	1175	
		:	:	:	:	:		
Qy	629	QLRYASALS	HIAFNALDND	GDGFQDNKVVVKT	HNSYALADGSHFV	----	676	
		:	:	:	:	:		
Db	1176	NEEISQ	-----	LNDEITSTQ	ENESIKKKND	-----LEGEVKAMKSTSE	QSNLKK 1222	
		:	:	:	:	:		
Qy	677	DPKLSSTI	VDLEK	-----	ILKULEGLDYTL	-----	ROTTQTFQFRDMTT 716	
		:	:	:	:	:		
Db	1223	SEIDALNLOI	KELKQXNET	NEASLLESIK	SEVETVKIKELQD	CNCFKEVSELEDKL	1282	
		:	:	:	:	:		
Qy	717	AGKALLSKN	LRQCEKQFLQ	-----	EAOFFGRV	DLDKATAKAKALVTKAN	--- 765	
		:	:	:	:	:		
Db	1283	ASBDKNSKY	LEQEKESKI	KEELDAKTEL	KIQLEITNLSKAK	KBSESLSRUKTTS	1342	
		:	:	:	:	:		
Qy	766	--KNG-	QLLERSINK	AVLAYNNSA	IKKANVKRLEK	ELDLLTLGVEGKPLAQATMVQGVY	822	
		:	:	:	:	:		
Db	1343	ERKNAERQ	LEKLKNE	--IQIKNOA	-----	PEKERKLLN	-----EGSSTITQ	1381
		:	:	:	:	:		
Qy	823	LLKTPPLP	PEYIYGLNV	VPDKSGKL	-----	IYALDMSDDTIGEGOKDAYGNPILNV	DEBNE 877	
		:	:	:	:	:		
Db	1382	-----	EYSEKINTLE	DELTRLO	QENELKAKEIDNTR	SELEKVSLSNDELLEEKQ	N- 1431	
		:	:	:	:	:		
Qy	878	GYHALAVAT	IADVEGLD	KITILNSKL	SQLTISI	-----	QVPTAAHYRAGIFOAIQNAAAE 932	
		:	:	:	:	:		
Db	1432	-----	TIKSLOD	-EILSYKOKI	TENDEKLLSIEDR	NKDELSLEKEQ	-----LRAAQSKAK 1481	
		:	:	:	:	:		
Qy	933	AEQLLPKPG	THSEKSS	SESA	-----	NSKD--	RGLOSNPKTN 967	
		:	:	:	:	:		
Db	1482	VEBGLK	-----	LEBESSKERA	ELEKSKEMMKKLESTIES	N	1517	
		:	:	:	:	:		

RESULT 94

Q9V624	PRELIMINARY;	PRT; 2346 AA.
ID Q9V624		
AC Q9V624;		
DT 01-MAY-2000	(TREMBlrel. 13, Created)	
DT 01-MAY-2000	(TREMBlrel. 13, Last sequence update)	
DT 01-MAR-2004	(TREMBlrel. 26, Last annotation update)	

DE	CG8274-PA.	;	ORFNames=CG8274;	
GN	Name=Mtor.			
OS	Drosophila melanogaster		(Fruit fly).	
OC	Drosophila melanogaster			
OC	Eukaryota;	Metazoa;	Arthropoda;	Hexapoda; Insecta; Pterygota;
OC	Neoptera;	Endopterygota;	Diptera;	Brachycera; Muscomorpha;
OC	Ephydroidea;	Drosophilidae;	Drosophila.	
OX	NCBI TaxID=7227;			

[1]

RP SQUENCE FROM N.A.
RX MEDLINE=20190606; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.P., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz-C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckel A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Swirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 RL melanogaster euchromatic genome sequence."
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).

RN SEQUENCE FROM N.A.
 RP MEDLINE=22426070; PubMed=12537573;
 RX Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskaas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celisniker S.E.;
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 RL a genomics perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 [3]

[4]
RN
RP
RX
RX
RA
RA
RA
RA
RA
RA
RA
RA
RA
RA
RA
RT
RT
RL

SEQUENCE FROM N.A.
MEDLINE=22426069; PubMed=12537572;
Misa S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,
Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
Lewis S.E.;
"Annotation of the *Drosophila melanogaster* euchromatic genome: a
systematic review";
RT Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RL

RN	
RP	SEQUENCE FROM N.A.
RG	FlyBase;
RL	Submitted (SPP-2002) to the ENBL/GenBank/DBJ databases.

RN	[6]	
RP	SEQUENCE FROM N.A.	
RG	FlyBase;	
RL	Submitted (MAR-2004) to the ENBL/GenBank/DBJ databases.	
DR	ENBL; AEO03824; AAF58615.1; -	
DR	IntAct; Q9V624; -	
DR	FlyBase; FBgn0013756; Mtor.	
DR	GO; GO:0005643; C:nuclear pore; IDA.	
DR	GO; GO:0005634; C:nucleus; IDA.	
SQ	SEQUENCE 2346 AA; 262351 MW; C618FD3F9C350B21 CRC64;	

Query Match	3.9%	Score 198.5;	DB 2;	Length 2346;
Best Local Similarity	18.2%	Pred. No. 5.1;		
Matches 218;	Conservative 181;	Mismatches 396;	Indels 403;	Gaps 45;

Qy	6	KTVALT-----TTTGVVTHNQEVSLVKPEILKQTQASSISGADYASSGSKL	56
		: : : : : : : : : : : :	
Db	225	KTESLKLMOEQYBQAVKTI GELTSKTE---MQNDTAFKQNGQATEYVGKTKKELDAKEKL	281
Qy	57	KINETGPDVDTVDL-----SPKRTTPPKIKONLAKGPR-----BOELKAVTENTSESK	107


```
RESULT 96
Q862A8 PRELIMINARY; PRT; 1169 AA.
AC Q862A8;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Kinesin.
GN Name=KLP3;
OS Cochliobolus heterostrophus (Drechslera maydis).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
OX NCBI_TaxID=5016;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22627967; PubMed=12742059; DOI=10.1016/S1087-1845(03)00022-7;
RA Schoch C.L., Aist J.R., Yoder O.C., Gillian Turgeon B.;
RT "A complete inventory of fungal kinesins in representative filamentous
ascomycetes.";
RL Fungal Genet. Biol. 39:1-15(2003).
DR EMBL; AV230428; AAO59290.1; -.
DR HSSP; P17119; 1FPT.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
DR ATP-binding; Microtubule; Motor protein.
KW SEQUENCE 1169 AA; 129138 MW; 783AF904F844DDCE CRC64;
SQ
Query Match 3.9%; Score 197.5; DB 2; Length 1169;
Best Local Similarity 19.9%; Pred. No. 2.1;
Matches 198; Conservative 155; Mismatches 344; Indels 297; Gaps 48;
QY 45 ADYABSSGK-SKLKNETSGPVDVTVDLFSDKRTTPEKIDN-----LAGPREQ 94
DB 164 ADBIENSVCSPFEL-----YNEELRLSLPDDATKLIYDNDSSKSGSTTLVQGMEEC 217
QY 95 ELKAVTENTESKQITSGSQLESLSNKTVPSTNWEICDFITKGNLTVLGSKGV 154
DB 218 HLKTASEGI---KLLANGSHKQVATKCNLDSSRSHVFTIVYIKR-----TTEGQ 268
QY 155 EKLQTDHLVPSQADGTQLIQVAFPTPDKTAIAEYTSRAG-ENGEISQLDVGKE 213
DB 269 EYLS-----AGKLNLDLAG-----SENIQRSGAENKRAAEAGLINKS 306
QY 214 IINEGVNSVLLKKVTIP---TGKHTIGODA-----FVDKNIAEVNLPESLETIS 262
DB 307 LTLGRVNALVERSHPIYRESKLTLLQDSLGGRTKTCIATLSPAKSNLEETISTL- 365
QY 263 DYAPAHALK---QIDLPNLKAIGELAFFDNOITGKLSLPRLMLRAEAFKSNHIXTI 319
DB 366 DYAPFAKIRKPNQVQNPINKTL--LKEYTWEIE---KLKSELATQR-----NGVILT 416
QY 320 EFRGNLSKVIQSGFQNDLSQJLMPDGLKIESEAFPTGNPGDDHNNRVVLMTKSGRNP 379
DB 417 QENYDEITTISES-----RRILSEQRERLET-----NEVNLRTKV-EDL 455
QY 380 SGLATENTYVNDKSLWQESPIDYTKWLEEDFTYQKNSVTGFSNKGKQKVRN---KNLE 437
DB 456 FKLTATKFTQLKKDNF--QTQALDGTGKILE-----KTEIV-----LNHTRQNLDDETE 502
QY 438 IPKQINGV--TITEIGDANFRNVDFQNTLRKYDLEEVKLPSTIRKIGAFQSNLKSFP 495
DB 503 LRKAHQKTEELAEVGRDMMSTL-----GKTTSAID-----GLRSKIKR--KSELQSQNRNN 553
QY 496 BASDDLEBIKEGAFMNNRIETLELKDVLVTIGDAAFINHIYAIIVLPESVQEGISAPRQ 555
Db 554 NSS-QTVQVVDTRLVVEDRIE--EFHQEQEQLMDA-----ISERMQTFVRDELEK 599
QY 556 NGANNLIIPWGSKVTLGEMAFLSNPLEHLDLSEQQLTEIPVQ-AFSDNALKEVLLPASL 614
DB 600 LGAS-----QSFLQEKMEAYQTS-----AEVNGQTAOARDHNMNEVL--BEI 639
QY 615 KTIREEAFKKNHLKQLEVASALSHTAFNALDNDGDEQFDNKKVVVKTHHNSYALADGHEF 674
DB 640 KTLREEVTKIGAGLDLSAAETISANIITELDA---FHTQV-----HSSYA-SLGRDF 690
QY 675 IVDPKLSTTVDLKILKLEGLDYSTLRTOTQTFQDMTTAGKALLSKSNLRQ----- 730
DB 691 -----KTTFDLVLK-----DLNDQQAENERLHQVVVEANTALTEAHKASEGQLAK 735
QY 731 ----EKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGQLLERSINKAVLAYNNSA 786
DB 736 LVDEKQKASDRQELLAQI-----TALMTANA----- 763
QY 787 IKKANVRLKELDLTLGLVEGKGFLAQATMVQGVYLLKTPLPPEYVIGLVNVPDKSGK 846
DB 764 --DAQEKELEERLSAVRGEIAAANTAPEAKQ-----SAYSEGVNAWSDKSRD 808
QY 847 LIYALDMS-DTIGEGQKDAYGNPI-----LNV----- 873
DB 809 ILAGVSKSRRTVTKIKISDFAVSVSCRFPTSHANVDQAATEHSTLIKETTTSVHSTVKT 868
QY 874 -----EDNEGYHA---LAVATLADYEGLDIKTI---LNSKLS 904
DB 869 VEAQWAHLDTQLQSLDDIVSRIRQNNNAHAHTTSLAALSSTVGASYSIGEHLASSFD 928
QY 905 QLTSTIRQVPTAAVHRAGIFQAIQNAARAEQLLP 938
DB 929 RVQSESEMSA--HTAGLKETLPLSAESDIRAP 960
RESULT 97
Q07380 PRELIMINARY; PRT; 1790 AA.
AC Q07380; P89892;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE US01 protein.
GN Name=US01;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloecker H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z74106; CAA98621.1; -.
DR PIR; S67593; S67593.
DR HSSP; P01106; INKP.
DR SGD; S00002216; US01.
DR GO; GO:0005625; C:soluble fraction; IDA.
DR GO; GO:0006888; P:ER to Golgi transport; IMP.
DR GO; GO:0006481; P:protein complex assembly; IDA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR02017; Spectrin.
DR InterPro; IPR006955; US01_p115_C.
DR InterPro; IPR006953; US01_p115_C.
DR Pfam; PF04871; US01_p115_C; 1.
DR Pfam; PF04869; US01_p115_head; 1.
SQ SEQUENCE 1790 AA; 206450 MW; 90062544F55A52EE CRC64;
Query Match 3.9%; Score 197.5; DB 2; Length 1790;
```



```
Db 1842 NRMEQISRDNETRQLNQDHESDKDEIIDKREMLEENLEENPNSS-----SDSLENP 1892
Qy 380 SGLATENTVNPDKSLWQSPEDYTKWLEE-----DFTYQKNSVTGFSNKGLOKVRNK 434
Db 1893 EG-----KKGKLEHTHSSEELDSVRDETYSK-----YKGIENQITET 1929
Qy 435 NLEIPKQHGVTITEIGDNAPRVDPNQNTLRKYDLEEVKLPSTIRKIGAFQSNLKS 494
Db 1930 EIESVEQD-----TNIPGNS-KETEVDNDRADMEBEKVKIKIITE----- 1971
Qy 495 PRASDDLEIEKEGAFNNRIETLEKDKLVTTGDAAAFHN-----HIY-----AIVL 541
Db 1972 -ESEELEISKDTT--TSHSEKPSIEEQSVNIQVDSKNEINVQIEKNVQNEQEGDPILL 2028
Qy 542 PESVQIEGRSAFRONGANNLIFMGSKVKTIGEMAPLSNRLE-----HLDLSEQKLTET 595
Db 2029 EQGNKVAILENOKNEYNPQSLSHKERTILLEVDLLEGGMDTNCNLTSELNKNKCSIQII 2088
Qy 596 PVQAFSDNALKEVLLPASLKTIRREAPKKNHLKQLEVASALSHIAFNALDDNDGDEQFDN 655
Db 2089 PRASNTDNKLN-----KDITE---NKDDFSEIEKSVGEIHENGKOLLNKESAESDDV 2137
Qy 656 KVVVKTHERNS-----YALADGEHPVIDDPDK-----LSSTIVDL 689
Db 2138 PVQNKIEHDSNAGVIDQYEYRTDYENEKNVAFVGLPGKPEBERSDVTTRGSPVSDAPE 2197
Qy 690 KILKLEIGLDYSTLRQTQTQFRDWTAGKALLSKSLNRQGEKQKFLQEAQFPLGRVDLD 749
Db 2198 K--KLUHEPDL-----SNEKNKELNVSKIEIRYVGLDIILG 2230
Qy 750 KAIAKAEKALVTKKAT-----KNGQLLERSINKAVLAYNNNSAIKANVKRLEKELDLTLG 805
Db 2231 IDSSESDSVTIYKRPDTFVENVHVLSDIKS---LFENEKTVGLNEQKETEKNIAGM 2287
Qy 806 VE--GKGPLAQATMGVGVLLKTPPLPYIYGLNVYFDKSKL-----IYALDMSDTIGRG 860
Db 2288 IDVLNKNKNDPQSVR-----LSTSL-----YHKGLED--DDQGRSLSTALYDKGLDDQHYL 2337
Qy 861 QKDAYGNPILNVDED-----NEGHYHALA-VATLADYE-----GLDICKILNLSKLSQ 905
Db 2338 HDNSNNNNNNINIKNEVQCKENGCRNENSVFVDTTLRDVKNQKHEETKLIENTKIDN 2397
Qy 906 LTSIRQVPTAAVHRAGIFAQIAQNAARQLL-----PKPGTHSEKSSSSSESANSKOR 958
Db 2398 TEVDGVDITSDNRSS-FISVDNYDEKKETIINEKKNDPKPSRKEEKDVKHDDSENDNN 2456
Qy 959 GLQSNPK 965
Db 2457 NLSSNIK 2463

RESULT 99
Q7P2P7 PRELIMINARY; PRT; 197 AA.
AC Q7P2P7;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
GN Name=FNV1342;
OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium
OC NCBI_TaxID=209882;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49256;
RA Karpatal V., Ivanova N., Anderson I., Reznik G., Bhattacharyya A.,
RA Gardner W.L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,
RA Haseikorn R., Overbeek R., Kyrpides N.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC !- CAUTION: The sequence shown here is derived from an
```

```
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABF01000055; EAA24104.1; --
KW Hypothetical protein.
SQ SEQUENCE 197 AA; 22444 MW; 3ED165BAEB414D32 CRC64;

Query Match 3.9%; Score 197; DB 2; Length 197;
Best Local Similarity 26.6%; Pred. No. 0.19;
Matches 54; Conservative 43; Mismatches 92; Indels 14; Gaps 4;

Qy 402 IDYTKWLESDFTYQKNSVTGFSNKGLOKVK--RNKNLEIPK-OHNGVTTTEIGDNAPRV 458
Db 1 MDQNIWEYDDFTFKGDELKGMTQKQKVKLEKTDLVIPELTPDGLPLKKIADNAF--- 57
Qy 459 DPQNTLRKYDLEEVKLPSTIRKIGAFQSNLKSFEASDDLEIEKEGAFNNRIETLE 518
Db 58 -----YRGLTSTVIVTNTVESIGYDAFGVCKLKEVKLPKALVNEGFAFRNKLTKVE 110
Qy 519 LKDKLVTTGDAAAFHNHIIYAIVLPSVQIEGRSAFRONGANNLIFMGSKVKTIGEMAFLS 578
Db 111 FGSKVKRLEPSSFAMNELAEVTFPETLEYIGASAFYKKNFETVSPKSVTK-IDMYAPRK 169
Qy 579 NRLEHLDLSEQKLTETIPVQAPS 601
Db 170 NNHKVEVANSVDLHKFAFEPT 192

RESULT 100
Q8GC27 PRELIMINARY; PRT; 897 AA.
AC Q8GC27;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Internalin B, i-InlB2 protein precursor.
GN Name=i-InlB2;
OS Listeria ivanovii.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC NCBI_TaxID=1638;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19119;
RA Dominguez-Bernal G., Gonzalez-Zorn B., Garrido P.,
RA Vazquez-Boland J.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ271621; CAC81246.1; --
DR HSSP; P25146; IO6V.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007092; LRR SDS22.
DR Pfam; PF00560; LRR 1; 15.
DR PRINTS; PR00019; LEURICHRPT.
KW Signal.
FT SIGNAL 1 33 Potential.
FT CHAIN 33 897 i-InlB2 protein.
SQ SEQUENCE 897 AA; 100397 MW; 4CE2608B29DBF314 CRC64;

Query Match 3.9%; Score 197; DB 2; Length 897;
Best Local Similarity 17.8%; Pred. No. 1.6;
Matches 164; Conservative 172; Mismatches 373; Indels 214; Gaps 36;

Qy 98 AVTENTESEKQITSGSQLEQSKESLSLNKTVFSTSNWEICDPITKGNLTVG-LSKSGVEK 156
Db 20 AMCVNTNLEIKVQRAAS-----ISNPMPIQIIPDTNLAEVMKSLRKNSTVDVVSQELDQ 75
Qy 157 LSQTDHLVLPQAADGTQLIQVAFPTPDKKTAETVTSRAGENGEISQLDVGKEIN 216
Db 76 VESINAMEKNIKSIQGLQHLNLTQIYLSDNQIQDISYLTNL---NKLSEIYLSGNQIKD 132
Qy 217 EGEVPSVLLKKVVTPTGYKIHGDQAFVDNKNIAEVNLPESLETISDYAFALHAKQIDL 276
Db 133 IGHLANLNKLEKI-----FLOGNQLTDINLPAGLSNLTJLVLSNNQIKDCN 179
Qy 277 PDNLKAGELAFFDQNIQTKLSLPQLMLRAERAFKSNHIKTIEPRGNSLVKVGESFQD 336
```

Db 180 LEKSKLENVYLOQNLTDISIAGLSNLMLDLSNNQIKGINOLANLNKLNELYLEG 236
QY 337 NDLSQLMLPDGLEKTESAFTGNPODDHYNRRVVLWTKSGKNPSGLATENTYVNPDKSLW 396
Db 237 NQLTDISVLAGLSNLKTL-----DLNNRI-----KDRTLSTLVNLENLW 278
QY 397 QESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRNKNLEIPKOHNGVYITEIGDNAFR 456
Db 279 NNNQLINT-----NHLSSLKLKLSFNGNRVTDISSVAKLTNLTELDCSEN 325
QY 457 NVDFONKTLRKDYLBVLEVPSTIRKIGAPAFOSNNLKSFEASDDLEETKEGAFMNRRET 516
Db 326 QVDNINSIAKLTNLTLGLT-----EGNKVKDLSPLAQLTNTLTGNFRONQIND 373
QY 517 LELKDKLVITIGDAAPHINHIYAIVLPESVQBEIGRSAPRONGANNLIFMGSKVKTIGENAF 576
Db 374 ISILEKLPNLSLAPDKNKVDVSLAKLPKLTYLIFNDNQVTNIDSLAKLPHLVGVDF 432
QY 577 LSNRLEHLDLSEKQKLTPIPVQAF---SDNALKEVLLPASLKTIREBAFKKNHLKQLEVA 633
Db 433 SGNKVSNT-----KALTNLTKLRLFNANGNCIQDIALRGLTQLEELKARNRIMDI--- 484
QY 634 SALSHIAFNALDDNDGBQDFDNV-----VVKTHNSVAL---ADGEHFIVDPD 679
Db 485 SPL--IWLNDLDELDSNOAFINRDFQVNVTFINIVKDTGTGLIAPNSSDKVTYNPP 542
QY 680 KLSSTIVBLEKILKLEGLDYSTLRQTQTQPR-----DMTTAGKALLSKSNLRQ 729
Db 543 NLTNLPTVKK--EVCYTFNQSVMLGQATSPSGQVIOFPFNKLDVVKSNKAITAYGRVKP 600
QY 730 GE-----KQ-----KFLQEA-----QFPLGRVDLDKAIKA 755
Db 601 GKTSTVWTQPCRTKDTKQVMVSAYTGKNLRLREAQTTSGIYYQFSVGGKTIGWVEAQA 660
QY 756 ---EKALVTKKAT-----KNGQ-----LLERSINKAVLAYNNSAIKKANVRLEK 797
Db 661 LGIFYQSSMEKATGARTVRGKEQHFKYKLPVIDAIDGGTLA-----KWKGRLLI 713
QY 798 ELDLITG-----LVBEGK-----PL---AQATMVQGVYLLKTPLPPEYYI 835
Db 714 QREVTIGKEKWLLOQGVGWKATNLTVLYNKPMGTGYAKVKLAKGVVWSNPYNITGVKV 773
QY 836 --GLNVYFDKSGKLIYALDMSD-----TIG--EG-----OKDAYGNPI 869
Db 774 VASLSRYTGKNLRLREAQTSNGLYYQVRVGKKTIGWVEAKNVAVFYKPIMEKKYKGRF 833
QY 870 LNVDEDNGYHALAVATLADYEG 892
Db 834 VKYKQKQAYYMPVADAAISRG 856

Search completed: August 28, 2005, 10:45:28

Job time : 236 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:58:12 ; Search time 48 Seconds
(without alignments)
2020.552 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 1008

Sequence: 1 MKKHLKTVALTLTTSVWTH.....LGYTSVALLSLTAIKKKY 1008

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 110 summaries

Database : PIR 79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	0.9	245	T51594	glutathione transf
2	9	0.9	907	JE0176	orphan G protein-c
3	8	0.8	116	T46320	hypothetical prote
4	8	0.8	253	T08175	chlorophyll a/b-bi
5	8	0.8	284	E71613	hypothetical prote
6	8	0.8	296	C95961	probable sugar upt
7	8	0.8	304	A87258	hypothetical prote
8	8	0.8	329	T37695	hypothetical prote
9	8	0.8	429	G97025	probable membrane
10	8	0.8	458	B84048	argininosuccinate
11	8	0.8	478	C72658	hypothetical prote
12	8	0.8	503	JQ2248	beta-amylase (EC 3
13	8	0.8	535	S00222	beta-amylase (EC 3
14	8	0.8	555	S71365	ovule development
15	8	0.8	681	JCS929	serine/arginine-ri
16	8	0.8	776	A87330	sensory box histid
17	8	0.8	862	T49593	hypothetical prote
18	8	0.8	880	SYBSVS	valine-tRNA ligase
19	8	0.8	1114	T30299	dyein heavy chain
20	8	0.8	1188	JC4889	phosphatidylinosit
21	8	0.8	1355	VGBE11	149K glycoprotein
22	8	0.8	1642	T08880	NMDA receptor-bind
23	7	0.7	66	D90841	hypothetical prote
24	7	0.7	66	C85699	hypothetical prote
25	7	0.7	77	A82086	hypothetical prote
26	7	0.7	82	A60457	clusterin - sheep
27	7	0.7	88	H69318	conserved hypothet
28	7	0.7	93	D69873	hypothetical prote
29	7	0.7	100	C64431	conserved hypothet

30	7	0.7	103	1	TRVKG1	terminal region re
31	7	0.7	104	1	GENSM1	matrix Gla protein
32	7	0.7	106	2	T05603	H+-exporting ATPas
33	7	0.7	106	2	T51826	hypothetical prote
34	7	0.7	110	2	T51825	H+-exporting ATPas
35	7	0.7	111	2	E84325	hypothetical prote
36	7	0.7	118	2	E95881	hypothetical prote
37	7	0.7	121	2	B96647	hypothetical prote
38	7	0.7	124	1	QVZEL	F11 protein - vacc
39	7	0.7	124	2	S28695	hypothetical prote
40	7	0.7	124	2	H71137	hypothetical prote
41	7	0.7	126	2	C95270	hypothetical prote
42	7	0.7	130	2	H81371	hypothetical prote
43	7	0.7	133	2	T42513	J5L 15.2K protein
44	7	0.7	133	2	T28520	hypothetical prote
45	7	0.7	133	2	H72160	hypothetical prote
46	7	0.7	133	2	S33096	M5L protein - vari
47	7	0.7	138	1	WNMS16	submandibular glan
48	7	0.7	138	2	A30305	hypothetical prote
49	7	0.7	139	2	T35860	hypothetical prote
50	7	0.7	142	2	A82517	ribonuclease P XF2
51	7	0.7	146	1	JC6173	single-stranded DN
52	7	0.7	146	1	HBHY	hemoglobin beta ch
53	7	0.7	147	2	S22336	hemoglobin beta ch
54	7	0.7	150	2	E83057	hypothetical prote
55	7	0.7	152	2	S36108	superoxide dismuta
56	7	0.7	152	2	T48772	hypothetical prote
57	7	0.7	160	2	E82382	transcription regu
58	7	0.7	161	2	I70177	dyein-like protei
59	7	0.7	172	2	S56905	hypothetical prote
60	7	0.7	173	2	T45501	hypothetical prote
61	7	0.7	175	2	T02557	hypothetical prote
62	7	0.7	180	2	H86805	hypothetical prote
63	7	0.7	182	2	T32619	hypothetical prote
64	7	0.7	185	2	A64438	hypothetical prote
65	7	0.7	187	2	T09390	21k protein precu
66	7	0.7	188	2	I70176	dyein-like protei
67	7	0.7	192	2	E90187	hAM1 protein (impo
68	7	0.7	195	2	E75462	hypothetical prote
69	7	0.7	196	1	F71246	probable indolepyr
70	7	0.7	197	2	T29084	3-isopropylmalate
71	7	0.7	197	2	T44358	hypothetical prote
72	7	0.7	200	2	A96504	protein F9C16.19 [
73	7	0.7	200	2	I70171	dyein-like protei
74	7	0.7	204	2	G64006	hypothetical prote
75	7	0.7	212	2	S48952	hypothetical prote
76	7	0.7	214	2	B84206	hypothetical prote
77	7	0.7	215	2	G95069	ABC transporter, A
78	7	0.7	215	2	E97937	hypothetical prote
79	7	0.7	219	2	C75637	hypothetical prote
80	7	0.7	220	2	A64479	DNA-(apurinic or a
81	7	0.7	220	2	JQ1985	avrPma1 protein -
82	7	0.7	220	2	JQ1985	avrPma1 protein -
83	7	0.7	228	2	H83639	hypothetical prote
84	7	0.7	230	2	C84606	probable RAS type
85	7	0.7	231	2	JQ1601	replication protei
86	7	0.7	238	2	C72086	sam-dependent meth
87	7	0.7	238	2	D86538	SAM-dependent meth
88	7	0.7	239	2	H64343	hypothetical prote
89	7	0.7	242	2	A86719	oxidoreductase yhg
90	7	0.7	247	2	A85688	hypothetical prote
91	7	0.7	249	2	A84653	hypothetical prote
92	7	0.7	249	2	AB2991	transcription regu
93	7	0.7	249	2	G98292	hypothetical trans
94	7	0.7	251	2	A86727	transcription regu
95	7	0.7	254	2	A96761	unknown protein li
96	7	0.7	257	1	WMAD84	early B1A 28K prot
97	7	0.7	257	2	AA1865	hypothetical prote
98	7	0.7	258	2	S36431	hypothetical prote
99	7	0.7	258	2	T36359	hypothetical prote
100	7	0.7	259	2	T19091	hypothetical prote
101	7	0.7	263	2	G84965	glycerol uptake fa
102	7	0.7	264	2	S76595	hypothetical prote

103 7 0.7 264 2 T23866 hypothetical prote
104 7 0.7 268 2 T15693 hypothetical prote
105 7 0.7 269 2 D69996 amino acid ABC tra
106 7 0.7 269 2 AE1368 amino acid ABC tra
107 7 0.7 269 2 AF1737 amino acid ABC tra
108 7 0.7 269 2 B69381 hypothetical prote
109 7 0.7 269 2 T22739 hypothetical prote
110 7 0.7 270 2 E69996 amino acid ABC tra

ALIGNMENTS

RESULT 1

T51594
Glutathione transferase (SC 2.5.1.18) AtGST 10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004
C:Accession: T51594
R:Dixon, D.P.; Cole, D.J.; Edwards, R.
Plant Physiol. 119, 1568, 1999
A:Title: Identification and cloning of AtGST 10 (Accession Nos. AJ131580 and AJ132398),
A:Reference number: Z25404
A:Accession: T51594
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-245 <DIX>
A:Cross-references: UNIPROT:Q9ZRT5; EMBL:AJ131580; PDB:CAA10457.1
A:Experimental source: variety Columbia; seedling; whole plant
C:Genetics:
A:Gene: gzt10
C:Keywords: transferase

Query Match 0.9%; Score 9; DB 2; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAAAEQLL 937

DB 133 AAAAAEQLL 141

RESULT 2

J90176
Orphan G protein-coupled receptor precursor - human
C:Species: Homo sapiens (man)
C>Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 09-Jul-2004
C:Accession: J90176
R:McDonald, T.; Wang, R.; Bailey, W.; Xie, G.; Chen, F.; Caskey, C.T.; Liu, Q.
Biochem. Biophys. Res. Commun. 247, 266-270, 1998
A:Title: Identification and cloning of an orphan G protein-coupled receptor of the glyco
A:Reference number: J90176; MUID:98308104; PMID:9642114
A:Accession: J90176
A:Molecule type: mRNA
A:Residues: 1-907 <MCD>
A:Cross-references: UNIPROT:O75473; GB:AF062006; NID:93366801; PDB:1AC28019.1; PDB:9336
C:Comment: This protein is a receptor for a novel class of glycoprotein ligands.
C:Genetics:
A:Gene: HG38
A:Map position: 12q22-23
F:1-21/Domain: signal sequence #status predicted <SIG>
F:562-583/Domain: transmembrane #status predicted <TM1>
F:594-616/Domain: transmembrane #status predicted <TM2>
F:639-660/Domain: transmembrane #status predicted <TM3>
F:681-701/Domain: transmembrane #status predicted <TM4>
F:725-744/Domain: transmembrane #status predicted <TM5>
F:768-791/Domain: transmembrane #status predicted <TM6>
F:803-824/Domain: transmembrane #status predicted <TM7>

Query Match 0.9%; Score 9; DB 2; Length 907;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVOAF 600
DB 174 LTEIPVOAF 182

RESULT 3

T46320
Hypothetical protein DKFP434H0512.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T46320
R:Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23035
A:Accession: T46320
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-116 <AAA>
A:Cross-references: UNIPROT:Q9NSZ8; EMBL:AL137620
A:Experimental source: adult testis; clone DKFP434H0512
C:Genetics:
A:Note: DKFP434H0512.1

Query Match 0.8%; Score 8; DB 2; Length 116;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 930 AAAAAEQLL 937

DB 83 AAAAAEQLL 90

RESULT 4

T08175
Chlorophyll a/b-binding protein homolog LI818r-1 precursor - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T08175
R:Guertin, M.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z16375
A:Accession: T08175
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-253 <GUR>
A:Cross-references: UNIPROT:P93664; EMBL:X95326
A:Experimental source: strain CC620
C:Genetics:
A:Gene: LI818r-1
C:Superfamily: chlorophyll a/b-binding protein

Query Match 0.8%; Score 8; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 LPDNLKAI 283

DB 246 LPDNLKAI 253

RESULT 5

E71613
Hypothetical protein PF0490c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C:Accession: E71613
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.G.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: E71613
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-284 <GAR>
 A:Cross-references: UNIPROT:O96191; GB:AE001398; GB:AE001362; NID:g3845197; PIDN:AA7188
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0490C

Query Match 0.8%; Score 8; DB 2; Length 284;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 EVNLPESL 258
 |||||
 Db 269 EVNLPESL 276

RESULT 6.
 C95961
 probable sugar uptake ABC transporter periplasmic solute-binding protein precursor Smb21
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
 C:Accession: C95961

R.Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: C95961
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-296 <KUR>
 A:Cross-references: UNIPROT:Q92666; GB:AL591985; PIDN:CAC49355.1; PID:g15140841; GSPDB:G
 A:Experimental source: strain 1021, megaplasmid pSymb
 R.Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation
 C:Genetics:
 A:Gene: Smb21377
 A:Genome: plasmid

Query Match 0.8%; Score 8; DB 2; Length 296;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 ATAKAEKA 758
 |||||
 Db 98 ATAKAEKA 105

RESULT 7.
 A87258
 hypothetical protein CC0074 [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
 C:Accession: A87258
 R.Nierman, W.C.; Feildblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: A87258
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-304 <STO>
 A:Cross-references: UNIPROT:Q9ABZ6; GB:AE005673; NID:g13421171; PIDN:AAK22061.1; GSPDB:G
 C:Genetics:
 A:Gene: CC0074

Query Match 0.8%; Score 8; DB 2; Length 304;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 881 ALAVATLA 888
 |||||
 Db 131 ALAVATLA 138

RESULT 8
 T37695
 hypothetical protein SPAC14C4.10c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C:Accession: T37695
 R.Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z21737
 A:Accession: T37695
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-329 <DEV>
 A:Cross-references: UNIPROT:O13717; EMBL:Z98596; PIDN:CAB11203.1; GSPDB:GN00066; SPDB:SF

A:Experimental source: strain 972h-; cosmid c14C4
 C:Genetics:
 A:Gene: SPDB:SPAC14C4.10c
 A:Map position: 1
 A:introns: 140/1

Query Match 0.8%; Score 8; DB 2; Length 329;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 798 ELDLITGL 805
 |||||
 Db 4 ELDLITGL 11

RESULT 9
 G97025
 probable membrane protein [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
 C:Accession: G97025
 R.Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: G97025
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-429 <KUR>

A:Cross-references: UNIPROT:Q97KAL; GB:AE001437; PIDN:AAK78994.1; PID:g15023928; GSPDB:G
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC1018

Query Match 0.8%; Score 8; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 841 FDKSGKLI 848
 |||||
 Db 239 FDKSGKLI 246

RESULT 10
 B84048
 argininosuccinate lyase argH [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004

C:Accession: B84048
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 C:Accession: B84048
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-458 <STO>
 A:Cross-references: UNIPROT:Q9K821; GB:AP001518; GB:BA000004; NID:gl10175792; PIDN:BA8069
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: argH
 C:Superfamily: argininosuccinate lyase

Query Match 0.8%; Score 8; DB 2; Length 458;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 KAIKAEK 757
 DB 447 KAIKAEK 454

RESULT 11
 C72658
 hypothetical protein APE0691 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
 C:Accession: C72658
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
 DNA Res. 6, 83-101, 1999
 A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
 A:Reference number: A72450; MUID:99310339; PMID:10382966
 A:Accession: C72658
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-478 <RAW>
 A:Cross-references: UNIPROT:Q9YE79; DDBJ:AP000060; NID:95104188; PIDN:BAA79667.1; PID:dl
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE0691
 C:Superfamily: Aeropyrum pernix hypothetical protein APE0691

Query Match 0.8%; Score 8; DB 2; Length 478;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 LKVI GEAS 333
 DB 406 LKVI GEAS 413

RESULT 12
 JQ2248
 beta-amylase (EC 3.2.1.2) - rye
 N:Alternate names: 1,4-alpha-D-glucan maltohydrolase
 C:Species: Secale cereale (rye)
 C:Date: 03-May-1994 #sequence_revision 25-Oct-1996 #text_change 09-Jul-2004
 C:Accession: JQ2248
 R:Sadowski, J.; Rorat, T.; Cooke, R.; Delseny, M.
 Plant Physiol. 102, 315-316, 1993
 A>Title: Nucleotide sequence of a cDNA clone encoding ubiquitous beta-amylase in rye (Se
 A:Reference number: JQ2248; MUID:94151427; PMID:8108499
 A:Accession: JQ2248
 A:Molecule type: mRNA
 A:Residues: 1-503 <SAD>
 A:Cross-references: UNIPROT:Q08335; GB:Z11772; NID:9393449; PIDN:CAA77817.1; PID:G93450
 C:Function:
 A>Description: hydrolyzes 1,4-glycosidic linkages of starch, removing maltose units succ
 C:Superfamily: beta-amylase
 C:Keywords: glycoprotein; glycosidase; hydrolase; monomer; polysaccharide degradation

F:2-503/Product: beta-amylase #status predicted <MAT>
 F:184,378/Active site: Glu #status predicted
 F:249,338/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.8%; Score 8; DB 1; Length 503;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
 DB 55 GLVEGKGP 62

RESULT 13
 S00222
 beta-amylase (EC 3.2.1.2) - barley
 N:Alternate names: 1,4-alpha-D-glucan maltohydrolase
 C:Species: Hordeum vulgare (barley)
 C:Date: 30-Jun-1989 #sequence_revision 25-Oct-1996 #text_change 09-Jul-2004
 C:Accession: S00222; JX0299; PC2019
 R:Kreiss, M.; Williamson, M.; Buxton, B.; Pywell, J.; Hejgaard, J.; Svendsen, I.
 Eur. J. Biochem. 169, 517-525, 1987
 A>Title: Primary structure and differential expression of beta-amylase in normal and mut
 A:Reference number: S00222; MUID:88082785; PMID:2446870
 A:Accession: S00222
 A:Molecule type: mRNA
 A:Residues: 1-535 <KRR>
 A:Cross-references: UNIPROT:P16098; EMBL:X52321; NID:gl8917; PIDN:CAA36556.1; PID:gl8918
 A>Note: parts of this sequence, including the amino end of the mature protein, were dete
 R:Yoshigi, N.; Okada, Y.; Sahata, H.; Koshino, S.
 J. Biochem. 115, 47-51, 1994
 A>Title: PCR cloning and sequencing of the beta-amylase cDNA from barley.
 A:Reference number: JX0299; MUID:94245663; PMID:8188635
 A:Accession: JX0299
 A:Molecule type: mRNA
 A:Residues: 1-232, 'A', 234-346, 'S', 348-526, 'M', 528-535 <YOS>
 A:Cross-references: GB:D21349; NID:9415599; PIDN:BAA04815.1; PID:g464145
 A:Accession: PC2019
 A:Molecule type: protein
 A:Residues: 129-141; 152-162; 384-394; 440-457 <YO2>
 C:Genetics:
 A:Map position: 4
 C:Function:
 A>Description: hydrolyzes 1,4-glycosidic linkages of starch, removing maltose units succ
 C:Superfamily: beta-amylase
 C:Keywords: glycosidase; hydrolase; monomer; polysaccharide degradation; seed; tandem re
 F:2-535/Product: beta-amylase #status predicted <MAT>
 F:488-531/Region: 11-residue repeats
 F:184,378/Active site: Glu #status predicted

Query Match 0.8%; Score 8; DB 1; Length 535;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
 DB 55 GLVEGKGP 62

RESULT 14
 S71365
 ovule development protein aintegumenta - Arabidopsis thaliana
 N:Alternate names: AP2 domain-containing protein; protein T28119.30
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004
 C:Accession: S71365; T06022
 R:Klucher, K.
 submitted to the EMBL Data Library, November 1995
 A:Reference number: S71365
 A:Accession: S71365
 A:Molecule type: mRNA
 A:Residues: 1-555 <KLU>
 A:Cross-references: UNIPROT:Q42462; EMBL:U40256; NID:gl209098; PID:gl209099

R.; Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15484
A:Accession: T06022
A:Molecule type: DNA
A:Residues: 1-555 <REV>
A:Cross-references: EMBL:AL035709; GSPDB:GNO0062; ATSP:T28119.30
A:Experimental source: cultivar Columbia; BAC clone T28119

C:Genetics:
A:Gene: ATSP:T28119.30; ANT
A:Map position: 4
A:Introns: 111/1; 289/2; 317/1; 320/1; 349/3; 391/2; 417/1
Query Match 0.8%; Score 8; DB 2; Length 555;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTL 12
Db 536 LKTVALTL 543
Query Match 0.8%; Score 8; DB 2; Length 555;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15
JCS929
serine/arginine-rich protein-specific kinase (EC 2.-.-.-) 2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
A:Accession: JCS929

R.; Kuroyanagi, N.; Onogi, H.; Wakabayashi, T.; Hagiwara, M.
Biochem. Biophys. Res. Commun. 242, 357-364, 1998
A:Title: Novel SR-protein-specific kinase, SRPK2, disassembles nuclear speckles.
A:Reference number: JCS929; MUID:98113357; PMID:9446799
A:Accession: JCS929

A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-681 <KOR>
A:Cross-references: UNIPROT:Q8VCD9; UNIPROT:O54781
A:Experimental source: brain
C:Comment: This enzyme regulates the disassembly of the serine/arginine-rich proteins in a
C:Keywords: transferase

Query Match 0.8%; Score 8; DB 2; Length 681;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSE 951
Db 5 SEKSSSSE 12

RESULT 16
A87330
sensory box histidine kinase/response regulator [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
A:Accession: A87330

R.; Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gunn, M.L.; Haft, D.H.; Kolon
n, J.; Esmailova, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87330

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-776 <STO>
A:Cross-references: UNIPROT:Q9AAB9; GB:AE005673; NID:g13421867; PIDN:AAK22637.1; GSPDB:G
C:Genetics:
A:Gene: CC0652

Query Match 0.8%; Score 8; DB 2; Length 776;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 716 TAGKALLS 723
Db 455 TAGKALLS 462

RESULT 17

T49593

hypothetical protein B3E4.60 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004

A:Accession: T49593

R.; Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49593

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-862 <SCH>

A:Cross-references: UNIPROT:Q9PSX6; EMBL:AL355931; GSPDB:GNO0116; NCSP:B3E4.60

A:Experimental source: BAC clone B3E4; strain OR74A

C:Genetics:

A:Gene: NCSP:B3E4.60

A:Map position: 6

A:Introns: 181/2; 764/1

C:Superfamily: Saccharomyces cerevisiae probable membrane protein YOR165w

Query Match 0.8%; Score 8; DB 2; Length 862;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 KSFEASDD 500

Db 573 KSFEASDD 580

RESULT 18

SYBSVS

valine-tRNA ligase (EC 6.1.1.9) - Bacillus stearothermophilus

N:Alternate names: valyl-tRNA synthetase

C:Species: Bacillus stearothermophilus

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004

A:Accession: A26738

R.; Borgford, T.J.; Brand, N.J.; Gray, T.E.; Fersht, A.R.

Biochemistry 26, 2480-2486, 1987

A:Title: The valyl-tRNA synthetase from Bacillus stearothermophilus has considerable seq

A:Reference number: A26738; MUID:87271643; PMID:3300774

A:Accession: A26738

A:Molecule type: DNA

A:Residues: 1-880 <BOR>

A:Cross-references: UNIPROT:P11931; GB:M16318

C:Genetics:

A:Gene: valS

C:Superfamily: valine-tRNA ligase

C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

Query Match 0.8%; Score 8; DB 1; Length 880;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 KRLEKELD 800

Db 820 KRLEKELD 827

RESULT 19

T30299

dynein heavy chain isotype 7A - sea urchin (Tripneustes gratilla) (fragment)

C:Species: Tripneustes gratilla

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004

A:Accession: T30299

R.; Gibbons, B.H.; Arai, D.J.; Tang, W.J.; Hays, T.S.; Gibbons, I.R.

Mol. Biol. Cell 5, 57-70, 1994

A:Title: Phylogeny and expression of axonemal and cytoplasmic dynein genes in sea urchin
 A:Reference number: Z20814; MUID:94243035; PMID:8186465

A:Accession: T30299

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1114 <GB>

A:Cross-references: UNIPROT:Q27811; EMBL:U03978; NID:G433384; PID:G433385; PIDN:AAA63592

C:Genetics:

A:Gene: DYH7A

C:Superfamily: dynein heavy chain, ciliary

Query Match 0.8%; Score 8; DB 2; Length 1114;

Best Local Similarity 100.0%; Pred. No. 68;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 275 DLPDLNKA 282

Db 152 DLPDLNKA 159

RESULT 20

JC4889

phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase (EC 3.1.3.-) - human

N:Alternate names: hps1CN5phosphatase

C:Species: Homo sapiens (man)

C>Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 09-Jul-2004

A:Accession: JC4889; PC4187

R:Drayner, A.L.; Peesee, X.; De Smedt, F.; Woischolski, R.; Parker, P.; Erneux, C.

Biochem. Biophys. Res. Commun. 225, 243-249, 1996

A:Title: Cloning and expression of a human placenta inositol 1,3,4,5-tetrakisphosphate

A:Reference number: JC4889; MUID:96332436; PMID:8769125

A:Contents: Placenta

A:Accession: JC4889

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1188 <DRA>

A:Cross-references: UNIPROT:Q92656; EMBL:X98429; NID:G1495455; PIDN:CAA67071.1; PID:e249

A:Accession: PC4187

A:Molecule type: protein

A:Residues: 582-592; 668-675 <DR2>

C:Comment: This enzyme plays a function in the phosphatidylinositol 3-kinase and isositol

sequence motifs show that this enzyme interacts with various proteins in signal transdu

C:Comment: Inositol 1,3,4,5-tetrakisphosphate is also a substrate for this enzyme.

C:Keywords: phosphoric monoester hydrolase

F:5-101/Domain: SH2 homology <SH2>

F:380/Binding site: substrate (Arg) #status predicted

F:671/Active site: Cys #status predicted

Query Match 0.8%; Score 8; DB 2; Length 1188;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 SKLQLTS 908

Db 262 SKLQLTS 269

RESULT 21

VGBE11

149K glycoprotein - ictaluriid herpesvirus 1 (strain auburn 1)

C:Species: ictaluriid herpesvirus 1

A:Note: host Ictalurus punctatus (channel catfish)

C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004

C:Accession: B36791

R:Davidson, A.J.

submitted to GenBank, January 1992

A:Description: Channel catfish virus: a new type of herpesvirus.

A:Reference number: A36804

A:Accession: B36791

A:Molecule type: DNA

A:Residues: 1-1355 <DAV>

A:Cross-references: UNIPROT:Q00104; GB:M75136; NID:G331209; PIDN:AAA88149.1; PID:G331256

R:Davidson, A.J.

Virology 186, 9-14, 1992

A:Title: Channel catfish virus: a new type of herpesvirus.

A:Reference number: A39447; MUID:92087490; PMID:1727613

A:Contents: annotation

A:Note: neither amino acid nor nucleotide sequence is given

C:Genetics:

A:Gene: 46

C:Superfamily: ictaluriid herpesvirus 149K glycoprotein

C:Keywords: Glycoprotein

F:81.112,129,169,173,192,542,655,682,744,780,811,815,860,865,868,882,895,1195,1213,1225,

Query Match 0.8%; Score 8; DB 1; Length 1355;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 881 ALAVATLA 888

Db 962 ALAVATLA 969

RESULT 22

T08880

NMDA receptor-binding protein yotiao - human

C:Species: Homo sapiens (man)

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

A:Accession: T08880

R:Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.

J. Neurosci. 18, 2017-2027, 1998

A:Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts with

A:Reference number: 216511; MUID:98151389; PMID:9482789

A:Accession: T08880

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1642 <LIN>

A:Cross-references: EMBL:AF026245; NID:G2623067; PIDN:AAB86384.1; PID:G2623068

C:Genetics:

A:Map position: 7q21-22

C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscle

Query Match 0.8%; Score 8; DB 2; Length 1642;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKDKL 523

Db 357 TLELKDKL 364

RESULT 23

D90841

hypothetical protein ECs1700 [imported] - Escherichia coli (strain O157:H7, substrain RI

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C:Accession: D90841

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

Genes. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90841

A:Molecule type: DNA

A:Residues: 1-66 <HAY>

A:Cross-references: UNIPROT:Q8XDH8; GB:BA000007; PIDN:BA835123.1; PID:G13361165; GSPDB:B

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1700

Query Match 0.7%; Score 7; DB 2; Length 66;

Best Local Similarity 100.0%; Pred. No. 58;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 970 RHSAILP 976

Db 50 RHSAILP 56
|||||
RESULT 24
C85699
Hypothetical protein Z1967 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: C85699
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85699
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-66 <STO>
A:Cross-references: UNIPROT:Q8XDH8; GB:AE0051174; NID:G12514907; PIDN:AA656055.1; GSPDB:C
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1967

Query Match 0.7%; Score 7; DB 2; Length 66;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 970 RHSAILP 976
|||||
Db 50 RHSAILP 56
|||||

RESULT 25
A82086
Hypothetical protein VC2357 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: A82086
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: A82086
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-77 <HE1>
A:Cross-references: UNIPROT:Q9KPL0; GB:AE004306; GB:AE003852; NID:G9656924; PIDN:AAF9550
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2357
A:Map position: 1

Query Match 0.7%; Score 7; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 KLEKEL 799
|||||
Db 57 KLEKEL 63
|||||

RESULT 26
A60457
Clusterin - sheep (fragments)
N:Alterin - sheep (fragments)
C:Species: Ovis orientalis aries. Ovis ammon aries (domestic sheep)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 19-Oct-1995
C:Accession: A60457; S10671; S10672
R:Griswold, M.D.; Bishop, P.D.; Kim, K.H.; Ping, R.; Siiteri, J.E.; Morales, C.
Ann. N. Y. Acad. Sci. 564, 154-172, 1989

A:Title: Function of vitamin A in normal and synchronized seminiferous tubules.
A:Reference number: A60457; MUID:89372568; PMID:2672955
A:Accession: A60457
A:Molecule type: protein
A:Residues: 1-82 <GR1>
R:Tsuruta, J.K.; Wong, K.; Fritz, I.B.; Griswold, M.D.
Biochem. J. 268, 571-578, 1990
A:Title: Structural analysis of sulphated glycoprotein 2 from amino acid sequence. Relat
A:Reference number: S10671; MUID:90303244; PMID:2363694
A:Accession: S10671
A:Molecule type: protein
A:Residues: 4-26; 41-49 <TSU>
A:Accession: S10672
A:Molecule type: protein
A:Residues: 62-82 <TS2>
C:Superfamily: clusterin
C:Keywords: glycoprotein; testis

Query Match 0.7%; Score 7; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 603 NALKEVL 609
|||||
Db 26 NALKEVL 32
|||||

RESULT 27
H69318
conserved hypothetical protein AF0552 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: H69318
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glock, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: H69318
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-88 <KLE>
A:Cross-references: UNIPROT:O29699; GB:AE001066; GB:AE000782; NID:G2689389; PIDN:AA89068
C:Superfamily: probable sulfur carrier protein slr0821
F:88/Modified site: 1-thioglycine (Gly) #status predicted

Query Match 0.7%; Score 7; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 687 DLEKILK 693
|||||
Db 39 DLEKILK 45
|||||

RESULT 28
D69873
hypothetical protein ylan - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: D69873
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteri
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Siroz, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, M.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: D59873
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-93 <KUN>
A;Cross-references: UNIPROT:007638; GB:Z99111; GB:AL009126; NID:92633699; PIDN:CAB13357
A;Experimental source: strain 168
C;Genetics:
A;Gene: ylan

Query Match 0.7%; Score 7; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 689 EKILKLI 695
|||||
Db 22 EKILKLI 28

RESULT 29
C64431
conserved hypothetical protein MJ1052 - *Methanococcus jannaschii*
C;Species: *Methanococcus jannaschii*
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: C64431
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Reon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A>Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: C64431
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-100 <BUL>
A;Cross-references: UNIPROT:Q58452; GB:U67548; GB:L77117; NID:92826361; PIDN:AAB99055.1;
C;Genetics:
A;Map position: REV989804-989502
A;Start codon: GTG
C;Superfamily: *Methanococcus jannaschii* conserved hypothetical protein MJ1052

Query Match 0.7%; Score 7; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 280 LKAIGEL 286
|||||
Db 94 LKAIGEL 100

RESULT 30
TRVKG1
terminal region recognition factor 1 - Yeast (*Kluyveromyces marxianus* var. *lactis*) plasmid
N;Alternate names: DNA-binding protein TRF1
C;Species: *Kluyveromyces marxianus* var. *lactis*, *Candida spharctica*
A;Note: plasmid PK12, K2
C;Date: 30-Sep-1989 #sequence_revision 09-May-1997 #text_change 09-Jul-2004
A;Accession: S00969; S10336; S17251; A41628
R;Tommasino, M.; Ricci, S.; Galeotti, C.L.
Nucleic Acids Res. 16, 5863-5878, 1988
A>Title: Genome organization of the killer plasmid pGK12 from *Kluyveromyces lactis*.
A;Reference number: S00959; MUID:88289339; PMID:3041369
A;Accession: S00968
A;Molecule type: DNA
A;Residues: 1-103 <TOM>
A;Cross-references: UNIPROT:P05476; EMBL:X07776; NID:92868; PIDN:CAA30611.1; PID:g2878

R;Wilson, D.W.; Meacock, P.A.
Nucleic Acids Res. 16, 8097-8112, 1988
A>Title: Extracellular gene expression in yeast: evidence for a plasmid-encoded RNA polymerase.
A;Reference number: S10336; MUID:98335549; PMID:3138657
A;Accession: S10336
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-103 <WIL>
A;Cross-references: EMBL:X07946; NID:92883; PIDN:CAA30767.1; PID:g2884
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1988
R;McNeel, D.G.; Tamanoi, F.
Proc. Natl. Acad. Sci. U.S.A. 88, 11398-11402, 1991
A>Title: Terminal region recognition factor 1, a DNA-binding protein recognizing the invasin of *Yersinia enterocolitica*.
A;Reference number: A41628; MUID:92107957; PMID:1763054
A;Contents: annotation; purification and activity after expression in *E. coli*
R;Tommasino, M.
Yeast 7, 245-252, 1991
A>Title: Killer system of *Kluyveromyces lactis*: the open reading frame 10 of the pGK12 plasmid.
A;Reference number: S17251; MUID:91353078; PMID:1882549
A;Accession: S17251
A;Molecule type: DNA
A;Residues: 1-71; 83-99 <TOM>
A;Cross-references: PIDN:AAB19610.1; PID:g234312; PIDN:AAB19611.1; PID:g234313
C;Genetics:
A;Genome: plasmid
C;Function:
A;Description: DNA binding; may play a role in the protein-primed initiation of the replication of the plasmid.
C;Superfamily: DNA-binding protein TRF1
C;Keywords: DNA binding; DNA replication; plasmid replication

Query Match 0.7%; Score 7; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 999 LITAICK 1005
|||||
Db 9 LITAICK 15

RESULT 31
GEMSM1
matrix Gla protein precursor - mouse
N;Alternate names: MGP
C;Species: *Mus musculus* (house mouse)
C;Date: 07-Sep-1990 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
A;Accession: JQ0455; I55484
R;Ikeda, T.
submitted to JPIID, May 1990
A;Reference number: JQ0455
A;Accession: JQ0455
A;Molecule type: mRNA
A;Residues: 1-104 <IKE>
A;Cross-references: UNIPROT:P19788; DBJ:D00613; NID:9220478; PIDN:BAA00488.1; PID:g2204
A;Experimental source: strain C57BL/6, cell type osteoblast
R;Luo, G.; D'Souza, R.; Hogue, D.; Karsenty, G.
J. Bone Miner. Res. 10, 325-334, 1995
A>Title: The matrix Gla protein gene is a marker of the chondrogenesis cell lineage during endochondral ossification.
A;Reference number: I55484; MUID:95274434; PMID:7754814
A;Accession: I55484
A;Status: translation not shown; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-24, 1', 26-104 <LUO>
A;Cross-references: GB:S77350; NID:g998473
A;Experimental source: 129 SvEv
A;Note: the source is designated as mouse and classified as *Mus sp.* by NCBI
C;Comment: This vitamin K-dependent calcium binding protein is secreted by most known cells than osteocalcin.
C;Genetics:
A;Introns: 21/1; 32/1; 57/2
C;Superfamily: osteocalcin
C;Keywords: bone; calcium binding; carboxylglutamic acid; cartilage; phosphoprotein; vitamin K
P;1-19/Domain: signal sequence #status predicted <SIG>
P;20-104/Product: matrix Gla protein #status predicted <MAT>

F;21,60,67,71/Modified site: gamma-carboxyglutamic acid (Glu) #status predicted
F;22,25,28/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted
F;73-79/disulfide bonds: #status predicted

Query Match 0.7%; Score 7; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 881 ALAVATL 887
|||||||

Db 12 ALAVATL 18

RESULT 32

T05603

H+-exporting ATPase (EC 3.6.3.6) chain G, vacuolar - Arabidopsis thaliana
N:Alternate names: protein F9D16.180
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T05603
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.P.
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15419
A:Accession: T05603
A:Molecule type: DNA
A:Residues: 1-106 <BEV>
A:Cross-references: UNIPROT:O82629; EMBL:AL035394
A:Experimental source: cultivar Columbia; BAC clone F9D16
C:Genetics:
A:Gene: vag2

A:Map position: 4
A:Introns: 26/1; 59/3
A:Note: F9D16.180

C:Superfamily: yeast H+-transporting ATPase 13K chain
C:Keywords: ATP; hydrogen ion transport; hydrolase
Query Match 0.7%; Score 7; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 790 ANVKRL 796
|||||||

Db 66 ANVKRL 72

RESULT 33

T51826

Hypothetical protein vag2 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004
C:Accession: T51826
R:Rouquie, D.; Tournaire-Roux, C.; Szponarski, W.; Rossignol, M.; Dumas, P.
submitted to the EMBL Data Library, May 1998
A:Description: Cloning and expression of G subunits of vacuolar-type ATPase from plants.
A:Reference number: Z25473
A:Accession: T51826
A:Molecule type: mRNA
A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Residues: 1-106 <ROU>
A:Cross-references: UNIPROT:O82629; EMBL:AJ005902; PIDN:CAA06759.1
C:Genetics:
A:Gene: vag2

C:Superfamily: yeast H+-transporting ATPase 13K chain
Query Match 0.7%; Score 7; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 790 ANVKRL 796
|||||||

Db 66 ANVKRL 72

RESULT 34

T51825

H+-exporting ATPase (EC 3.6.3.6) chain G, vacuolar [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004
C:Accession: T51825
R:Rouquie, D.; Tournaire-Roux, C.; Szponarski, W.; Rossignol, M.; Dumas, P.
submitted to the EMBL Data Library, May 1998
A:Description: Cloning and expression of G subunits of vacuolar-type ATPase from plants
A:Reference number: Z25473
A:Accession: T51825
A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA
A:Residues: 1-110 <ROU>
A:Cross-references: UNIPROT:O82628; EMBL:AJ005901; PIDN:CAA06758.1
A:Experimental source: cultivar Columbia; leaves
C:Genetics:
A:Gene: vag1

C:Superfamily: yeast H+-transporting ATPase 13K chain
C:Keywords: ATP; hydrogen ion transport; hydrolase
Query Match 0.7%; Score 7; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 790 ANVKRL 796
|||||||

Db 70 ANVKRL 76

RESULT 35

E84325

Hypothetical protein Vng1734h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: E84325
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, E.; Leitchauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: E84325
A:Molecule type: DNA
A:Status: preliminary

A:Residues: 1-111 <STO>
A:Cross-references: UNIPROT:Q9HPA1; GB:AE004437; NID:g10581194; PIDN:AAG19969.1; GSPDB:G

C:Genetics:
A:Gene: VNG1734H

Query Match 0.7%; Score 7; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 882 LAVATLA 888
|||||||

Db 44 LAVATLA 50

RESULT 36

E95881

Hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSYMB
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C:Accession: E95881
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95881
A:Status: preliminary

A:Molecule type: DNA

A;Residues: 1-118 <KUR>
A;Cross-references: UNIPROT:Q92MW2; GB:AL591985; PIDN:CAC48717.1; PID:gl5140190; GSPDB:G
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibart, F.; Finan, T.M.; Long, S.R.; Fuhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbot, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
C;Contents: annotation
C;Genetics:
A;Gene: SMB20331
A;Genome: plasmid

Query Match 0.7%; Score 7; DB 2; Length 118;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 881 ALAVATL 887
|||||
Db 51 ALAVATL 57

RESULT 37
B96647
hypothetical protein F19K23.2 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: B96647
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: B96647
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-121 <STO>
A;Cross-references: UNIPROT:Q04574; GB:AE005173; NID:g2160134; PIDN:AAB60756.1; GSPDB:GN
C;Genetics:
A;Gene: F19K23.2
A;Map position: 1

Query Match 0.7%; Score 7; DB 2; Length 121;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 825 KTLPLP 831
|||||
Db 46 KTLPLP 52

RESULT 38
QOVZEL
Flt protein - vaccinia virus
C;Species: vaccinia virus
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 09-Jul-2004
C;Accession: B26216; L23092
R;Plucieniczak, A.; Schroeder, E.; Zettlmeisel, G.; Streeck, R.E.
Nucleic Acids Res. 13, 985-998, 1985
A;Title: Nucleotide sequence of a cluster of early and late genes in a conserved segment
A;Reference number: A23092; MUID:85215527; PMID:2987815
A;Accession: B26216
A;Molecule type: DNA
A;Residues: 1-124 <PLU>

A;Cross-references: UNIPROT:P07618; GB:X01978; GB:J02424; GB:J02425; GB:K02376; GB:ML521
A;Experimental source: strain WR
C;Comment: It is unknown whether this protein is synthesized before or after viral DNA r
C;Superfamily: vaccinia virus Flt protein

Query Match 0.7%; Score 7; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615
|||||
Db 52 LLPASLK 58

RESULT 39
S28695
hypothetical protein 13 - Agrobacterium tumefaciens plasmid pTi15955
C;Species: Agrobacterium tumefaciens
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: S28695
R;Barker, R.F.; Idler, K.B.; Thompson, D.V.; Kemp, J.D.
Plant Mol. Biol. 2, 335-350, 1983
A;Title: Nucleotide sequence of the T-DNA region from the Agrobacterium tumefaciens octo
A;Reference number: S28683
A;Accession: S28695
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-124 <BAR>
A;Cross-references: UNIPROT:Q44392; EMBL:X00493; NID:g39062; PIDN:CAA25175.1; PID:g39075
C;Genetics:
A;Genome: plasmid

Query Match 0.7%; Score 7; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 970 RHSAILP 976
|||||
Db 5 RHSAILP 11

RESULT 40
H71137
hypothetical protein PH0868 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C;Accession: H71137
R;Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kuehida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: H71137
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-124 <KAW>
A;Cross-references: UNIPROT:O58598; GB:AF000003; NID:g3236130; PIDN:BA29962.1; PID:g325
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0868

Query Match 0.7%; Score 7; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 994 VALLSLI 1000
|||||
Db 18 VALLSLI 24

RESULT 41
C95270

hypothetical protein Sma0128 [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
 C:Accession: C95270
 R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
 .; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9886, 2001
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
 A:Reference number: A95262; MUID:21396509; PMID:11481432
 A:Accession: C95270
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-126 <KUR>
 A:Cross-references: UNIPROT:Q930X7; GB:AE006469; PIDN:AAK64725.1; PID:g14523128; GSPDB:G
 A:Experimental source: strain 1021, megaplasmid pSymA
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: Sma0128
 A:Genome: plasmid

Query Match 0.7%; Score 7; DB 2; Length 126;
 Best Local Similarity 100.0%; Pred. No. 1.e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 929 AAAREAQ 935

Db 79 AAAREAQ 85

RESULT 42

H81371
 hypothetical protein Cj0971 [imported] - Campylobacter jejuni (strain NCTC 11168)
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
 C:Accession: H81371
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
 C.W.; Quail, M.; Rajadream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
 A:Reference number: A81250; MUID:20150912; PMID:10688204
 A:Accession: H81371
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-130 <PAR>
 A:Cross-references: UNIPROT:Q9PNW7; GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CB7322
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj0971
 C:Superfamily: Campylobacter jejuni hypothetical protein Cj0971

Query Match 0.7%; Score 7; DB 2; Length 130;
 Best Local Similarity 100.0%; Pred. No. 1.e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 NALKEVL 609

Db 99 NALKEVL 105

RESULT 43

H42513
 J5L 15.2K protein - vaccinia virus (strain Ankara, Copenhagen)
 C:Species: vaccinia virus
 A:Note: host Homo sapiens (man)
 C:Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 09-Jul-2004

C:Accession: H42513; T37365
 R:Johnson, G.P.
 submitted to GenBank, June 1990
 A:Reference number: A31172

A:Accession: H42513

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-133 <JOH>

A:Cross-references: UNIPROT:P21083

R:Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F.

submitted to the EMBL Data Library, March 1997

A:Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) strain

A:Reference number: Z20877

A:Accession: T37365

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-133 <ANT>

A:Cross-references: EMBL:U94848; PIDN:AB964445.1

A:Experimental source: strain Ankara

C:Genetics:

A:Note: MVA089L

C:Superfamily: vaccinia virus F11 protein

Query Match 0.7%; Score 7; DB 2; Length 133;
 Best Local Similarity 100.0%; Pred. No. 1.e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615

Db 52 LLPASLK 58

RESULT 44

T28520

hypothetical protein LSL - variola major virus

C:Species: variola major virus

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004

C:Accession: T28520

R:Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Aubin

Nature 366, 748-751, 1993

A:Title: Potential virulence determinants in terminal regions of variola smallpox virus

A:Reference number: Z20488; MUID:94088747; PMID:8264798

A:Accession: T28520

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-133 <MAS>

A:Cross-references: UNIPROT:Q85384; EMBL:L22579; NID:G623595; PIDN:AAA60830.1; PID:G4390

A:Experimental source: strain Bangladesh-1975

C:Superfamily: vaccinia virus F11 protein

Query Match 0.7%; Score 7; DB 2; Length 133;
 Best Local Similarity 100.0%; Pred. No. 1.e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615

Db 52 LLPASLK 58

RESULT 45

H72160

M5L protein - variola minor virus (strain Garcia-1966)

C:Species: variola minor virus

C:Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 09-Jul-2004

C:Accession: H72160

R:Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safronov, P.F.; Massung, R.F.; Lopar

submitted to GenBank, March 1998

A:Description: Analysis of the complete coding sequence of DNA of alastrim variola minor

A:Reference number: A72150

A:Accession: H72160

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-133 <SHC>

A;Cross-references: UNIPROT:P33055; GB:Y16780; NID:G5830555; PIDN:CAB54682.1; PID:G58306
 A;Experimental source: strain Garcia-1966
 C;Genetics: M5L
 C;Superfamily: vaccinia virus F11 protein

Query Match 0.7%; Score 7; DB 2; Length 133;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615
 |||||
 Db 52 LLPASLK 58

RESULT 46
 S33096
 J5L protein - variola virus
 N;Alternate names: L5L protein
 C;Species: variola virus
 C;Date: 02-Dec-1993 #sequence_revision 26-May-1995 #text_change 09-Jul-2004
 C;Accession: S33096; G36845
 R;Shchelkunov, S.N.; Blinov, V.M.; Totmenin, A.V.; Marennikova, S.S.; Kolykhalov, A.A.;
 dshaperidze, O.G.; Sandakhchiev, L.S.
 Virus Res. 27, 25-35, 1993
 A;Title: Nucleotide sequence analysis of variola virus HindIII M, L, I genome fragments.
 A;Reference number: S33069; MUID:93190624; PMID:8383392
 A;Accession: S33096
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-133 <SHC>
 A;Cross-references: UNIPROT:P33055; EMBL:X67119; NID:G62330; PIDN:CAA47581.1; PID:G62358
 A;Experimental source: strain India-1967, isolate Ind3
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992
 R;Blinov, V.M.
 submitted to GenBank, November 1992
 A;Reference number: A36859
 A;Accession: G36845
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-133 <BLI>
 A;Cross-references: GB:X69198; NID:G456758; PIDN:CAA49023.1; PID:G297262
 A;Experimental source: strain India-1967, esp. major, isolate Ind3
 C;Superfamily: vaccinia virus F11 protein

Query Match 0.7%; Score 7; DB 2; Length 133;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615
 |||||
 Db 52 LLPASLK 58

RESULT 47
 WMM516
 submandibular gland 16.5K protein - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 09-Jul-2004
 C;Accession: A03298
 R;Windas, J.D.; Mullins, J.J.; Beecroft, L.J.; George, H.; Meacock, P.A.; Williams, B.R.
 Nucleic Acids Res. 12, 1361-1376, 1984
 A;Title: Molecular cloning of cDNAs from androgen-independent mRNA species of DBA/2 mouse
 A;Reference number: A93503; MUID:84144035; PMID:6546617
 A;Accession: A03298
 A;Molecule type: mRNA
 A;Residues: 1-138 <WIN>
 A;Cross-references: UNIPROT:P02815; GB:X00349; NID:G51367; PIDN:CAA95098.1; PID:G51368
 C;Comment: this protein contains a hydrophobic amino-terminal sequence that is similar to
 C;Superfamily: submandibular gland 16.5K protein
 C;Keywords: glycoprotein; submandibular gland
 F;25,72,89,94/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.7%; Score 7; DB 1; Length 138;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 KSGVEKL 157
 |||||
 Db 125 KSGVEKL 131

RESULT 48
 A30305
 submandibular gland protein (spot 1) precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 22-Nov-1989 #sequence_revision 03-Aug-1992 #text_change 09-Jul-2004
 C;Accession: A30305
 R;Dickinson, D.P.; Mirels, L.; Tabak, L.A.; Gross, K.W.
 Mol. Biol. Evol. 6, 80-102, 1989
 A;Title: Rapid evolution of variants in a rodent multigene family encoding salivary protein
 A;Reference number: A30305; MUID:89158788; PMID:2921944
 A;Accession: A30305
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-138 <DIC>
 A;Cross-references: UNIPROT:P02815; GB:M33974; NID:G201044; PIDN:AAA40132.1; PID:G201045
 A;Note: the authors translated the codon AAC for residue 104 as Gln
 C;Superfamily: submandibular gland 16.5K protein

Query Match 0.7%; Score 7; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 KSGVEKL 157
 |||||
 Db 125 KSGVEKL 131

RESULT 49
 T35860
 hypothetical protein SC9B1.15c - Streptomyces coelicolor
 C;Species: Streptomyces coelicolor
 C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
 C;Accession: T35860
 R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
 submitted to the EMBL Data Library, April 1999
 A;Reference number: Z21591
 A;Accession: T35860
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-139 <SAU>
 A;Cross-references: UNIPROT:Q9X835; EMBL:AL049727; PIDN:CAB41561.1; GSPDB:GN00070; SCOD
 C;Genetics: A3(2)
 C;Gene: SCODB:SC9B1.15c
 C;Superfamily: Streptomyces coelicolor hypothetical protein SC1C2.20c

Query Match 0.7%; Score 7; DB 2; Length 139;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 801 LLTGLVE 807
 |||||
 Db 15 LLTGLVE 21

RESULT 50
 A82517
 ribonuclease P XP2781 [imported] - Xylella fastidiosa (strain 9a5c)
 C;Species: Xylella fastidiosa
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 16-Aug-2004
 C;Accession: A82517
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: For a complete list of authors see reference number A59328 below
A:Accession: A82517
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-142 <SIM>
A:Cross-references: UNIPROT:Q9P9U0; GB:AE004083; GB:AE003849; NID:99108022; PIDN:AAF8556
A:Experimental source: strain 9a5c
R:Simpsom, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tsuchioka, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XP2781
C:Superfamily: bacterial ribonuclease P, protein component
Query Match 0.7%; Score 7; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 993 SVALLSL 999
Db 109 SVALLSL 115
|||||
RESULT 51
JC6173
single-stranded DNA-binding protein 1 precursor, mitochondrial - African clawed frog
N:Alternate names: helix-desstabilizing protein; single-stranded DNA-binding protein s
C:Species: Xenopus laevis (African clawed frog)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: JC6173; S22300; S20262; S0116
R:Champagne, A.M.; Dufresne, C.; Viney, L.; Gueride, M.
Gene 184, 65-71, 1997
A:Title: Cloning, sequencing and expression of the two genes encoding the mitochondrial
A:Reference number: JC6173; MUID:97169147; PMID:9016954
A:Contents: oocyte
A:Accession: JC6173
A:Molecule type: mRNA
A:Residues: 1-146 <CHA>
A:Cross-references: UNIPROT:P09380; EMBL:X83673; NID:9620127; PIDN:CAA58647.1; PID:91890
R:Tiranti, V.; Barat-Gueride, M.; Bljl, J.; Didonato, S.; Zeviani, M.
Nucleic Acids Res. 19, 4291, 1991
A:Title: A full-length cDNA encoding a mitochondrial DNA-specific single-stranded DNA bi
A:Reference number: S22300; MUID:91334145; PMID:1870981
A:Accession: S22300
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-146 <TIR>
A:Cross-references: EMBL:X59285; NID:964898; PIDN:CAA41376.1; PID:964899
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1991
R:Ghrir, R.; Lecaer, J.P.; Dufrene, C.; Gueride, M.
Arch. Biochem. Biophys. 291, 395-400, 1991
A:Title: Primary structure of the two variants of Xenopus laevis mtSSB, a mitochondrial
A:Reference number: S20262; MUID:92061073; PMID:1952953
A:Accession: S20262
A:Molecule type: protein
A:Residues: 18-142 <GHR>
R:Mahongou, C.; Ghrir, R.; Lecaer, J.P.; Mignotte, B.; Barat-Gueride, M.
FEBS Lett. 235, 267-270, 1988
A:Title: The amino-terminal sequence of the Xenopus laevis mitochondrial SSB is homologo
A:Reference number: S01116; MUID:88296837; PMID:3042458
A:Accession: S01116

A:Molecule type: protein
A:Residues: 18-26, 'E', 28-41 <MAH>
C:Comment: This protein is essential for replication, repair of recombination. It is co
d, and it modulates the level of replication and transcription mediated by mitochondrial
C:Comment: This protein binds preferentially to single-stranded DNA.
C:Genetics:
A:Gene: ssb2
A:Genome: nuclear
A:Introns: 8/3; 27/1; 74/1; 104/2; 134/1
C:Superfamily: single-strand binding protein; single-stranded DNA-binding protein homol
C:Keywords: DNA replication; mitochondrion; oocyte; single-stranded DNA binding; tetran
F1-17/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F18-146/Product: single-stranded DNA-binding protein 1 #status predicted <MAT>
F144-129/Domain: single-stranded DNA-binding protein homology <SSD>
Query Match 0.7%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 771 LERSINK 777
Db 24 LERSINK 30
|||||
RESULT 52
HBHY
hemoglobin beta chain - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 12-Jul-2004
C:Accession: A02410
R:Braunitzer, G.; Schrank, B.; Stangl, A.; Wiesner, H.
J. Chem. Soc. Pak. 2, 1-7, 1980
A:Title: Respiration at high altitudes, phosphate-protein-interaction: sequence of the
A:Reference number: A92758
A:Accession: A02410
A:Molecule type: protein
A:Residues: 1-146 <BRA>
A:Cross-references: UNIPROT:P02094
C:Superfamily: globin; Globin homology
C:Keywords: blood; chromoprotein; erythrocyte; heme; iron; metalloprotein; oxygen carri
F13-146/Domain: globin homology <GHS>
F163/Binding site: oxygen (His) (distal axial ligand) #status predicted
F192/Binding site: heme iron (His) (proximal axial ligand) #status predicted
Query Match 0.7%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 755 AEKALVT 761
Db 6 AEKALVT 12
|||||
RESULT 53
S22336
hemoglobin beta chain, major - hamster
C:Species: Cricetinae gen. sp. (hamster)
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 12-Jul-2004
C:Accession: S22336; S36682
R:Lee, K.M.; Subar, M.; Li, H.; Boussios, T.
Biochim. Biophys. Acta 1130, 343-344, 1992
A:Title: Cloning of two adult hamster globin cDNAs (alpha and beta(major)).
A:Reference number: S22335; MUID:92223120; PMID:1562610
A:Accession: S22336
A:Molecule type: mRNA
A:Residues: 1-147 <LEE>
A:Cross-references: EMBL:X57030
R:Boussios, T.
submitted to the EMBL Data Library, December 1990
A:Reference number: S36682
A:Accession: S36682
A:Molecule type: mRNA
A:Residues: 1-115, 'I', 117-147 <BOU>

A;Cross-references: EMBL:X57030; NID:g49422; PIDN:CAA40346.1; PID:g49423
 C;Superfamily: globin; Globin homology
 C;Keywords: blood; chromoprotein; erythrocyte; heme; iron; metalloprotein; oxygen carrier
 F;4-147/Domain: globin homology <GLB>
 F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
 F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 0.7%; Score 7; DB 2; Length 147;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 755 AEKALVT 761
 |||||
 Db 7 AEKALVT 13

RESULT 54
 E83057
 hypothetical protein PA4697 [imported] - Pseudomonas aeruginosa (strain PA01)
 C;Species: Pseudomonas aeruginosa
 C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C;Accession: E83057
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A;Reference number: A82950; MUID:20437337; PMID:10984043
 A;Accession: E83057
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-150 <STO>
 A;Cross-references: UNIPROT:Q9HV99; GB:AB004884; GB:AE004091; NID:g9950954; PIDN:AAG0808
 A;Experimental source: strain PA01
 C;Genetics:
 A;Gene: PA4697

Query Match 0.7%; Score 7; DB 2; Length 150;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 752 IAKAEKA 758
 |||||
 Db 138 IAKAEKA 144

RESULT 55
 S36108
 superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - guinea pig
 C;Species: Cavia porcellus (Guinea pig)
 C;Date: 19-Mar-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
 C;Accession: S36108; S65796
 R;Wolf, B.; Reinecke, K.; Amann, K.D.; Brigelius-Flohe, R.; Flohe, L.
 Biol. Chem. Hoppe-Seyler 374, 641-649, 1993
 A;Title: Taxonomical classification of the guinea pig based on its Cu/Zn superoxide dis
 A;Reference number: S36108; MUID:94059414; PMID:8240718
 A;Accession: S36108
 A;Molecule type: protein
 A;Residues: 1-101; IL', 104-152 <WOL>
 A;Cross-references: UNIPROT:P33431
 R;Yuan, H.T.; Bingle, C.D.; Kelly, P.J.
 Biochim. Biophys. Acta 1305, 163-171, 1996
 A;Title: Differential patterns of antioxidant enzyme mRNA expression in guinea pig lung
 A;Reference number: S65793; MUID:96180320; PMID:8597602
 A;Accession: S65796
 A;Molecule type: mRNA
 A;Residues: 6-145 <YUA>
 A;Cross-references: EMBL:U39844
 C;Function:
 A;Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
 C;Superfamily: superoxide dismutase [Cu-Zn]
 C;Keywords: copper; metalloprotein; oxidoreductase; zinc
 F;45,47,62,119/Binding site: copper (His) #status predicted

F;56-145/Disulfide bonds: #status predicted
 F;62,70,79,82/Binding site: zinc (His, His, His, Asp) #status predicted
 F;142/Active site: Arg #status predicted

Query Match 0.7%; Score 7; DB 2; Length 152;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 803 TGLVEGK 809
 |||||
 Db 35 TGLVEGK 41

RESULT 56
 T48772
 hypothetical protein I3E11.200 [imported] - Neurospora crassa
 C;Species: Neurospora crassa
 C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 19-May-2000
 C;Accession: T48772
 R;Schulze, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, April 2000
 A;Reference number: Z24541
 A;Accession: T48772
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-152 <SCH>
 A;Cross-references: EMBL:ALJ53820; GSPDB:GN00112; NCSP:I3E11.200
 A;Experimental source: cosmid contig I3E11; strain 74
 C;Genetics:
 A;Gene: NCSP:I3E11.200
 A;Map position: 2
 C;Superfamily: Neurospora crassa hypothetical protein I3E11.200

Query Match 0.7%; Score 7; DB 2; Length 152;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 440 KOHNGVT 446
 |||||
 Db 66 KOHNGVT 72

RESULT 57
 E82382
 transcription regulator AsnC family VCA1068 [imported] - Vibrio cholerae (strain N16961
 C;Species: Vibrio cholerae
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
 C;Accession: E82382
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P
 l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A;Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A;Reference number: A82035; MUID:20406833; PMID:10952301
 A;Accession: E82382
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-160 <HEI>
 A;Cross-references: UNIPROT:Q9KKM4; GB:AB004432; GB:AE003853; NID:g9658509; PIDN:AAF9696
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor
 C;Genetics:
 A;Gene: VCA1068
 A;Map position: 2
 C;Superfamily: regulatory protein asnC

Query Match 0.7%; Score 7; DB 2; Length 160;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 792 VKRLEKE 798
 |||||
 Db 40 VKRLEKE 46


```

RESULT 58
I70177
dynein-like protein 7 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 09-Jul-2004
C:Accession: I70177
R.Tanaka, Y.; Zhang, Z.; Hirokawa, N.
J. Cell Sci. 108, 1883-1893, 1995
A:Title: Identification and molecular evolution of new dynein-like protein sequences in
A:Reference number: 155515; MUID:95386588; PMID:7657712
A:Accession: I70177
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-161 <RES>
A:Cross-references: UNIPROT:Q63170; GB:D26498; NID:g871910; PIDN:BAA05506.1; PID:g871911
C:Genetics:
A:Gene: DLP7

Query Match      0.7%; Score 7; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      276 LPDNLKA 282
DB      105 LPDNLKA 111

RESULT 59
S56905
Hypothetical protein YJL124c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein J0714
C:Species: Saccharomyces cerevisiae
C>Date: 05-May-1995 #sequence_revision 08-Sep-1995 #text_change 09-Jul-2004
C:Accession: S56905
R.Cziepluch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56891
A:Accession: S56905
A:Molecule type: DNA
A:Residues: 1-172 <Z1>
A:Cross-references: UNIPROT:P47017; EMBL:Z49399; NID:g1008319; PIDN:CAA89419.1; PID:g100
C:Genetics:
A:Gene: SGD:LSM1; MIPS:YJL124C
A:Cross-references: SGD:S0003660
A:Map position: 10L
C:Superfamily: hypothetical protein YJL124c

Query Match      0.7%; Score 7; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      952 SANSKDR 958
DB      2 SANSKDR 8

RESULT 60
T45501
Hypothetical protein yciD [imported] - Escherichia coli (fragment)
C:Species: Escherichia coli
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C:Accession: T45501
R.Milman, R.
submitted to the EMBL Data Library, April 1995
A:Description: Recombinational exchange among clonal populations.
A:Reference number: Z22991
A:Accession: T45501
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-173 <ML>
A:Cross-references: UNIPROT:P21364; EMBL:U24203; PIDN:AAB60129.1
A:Experimental source: strain ECOR 52

```

```

C:Genetics:
A>Note: yciD

Query Match      0.7%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      881 ALAVATL 887
DB      8 ALAVATL 14

RESULT 61
T02557
Hypothetical protein At2g32580 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T26B15.14
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 09-Jul-2004
C:Accession: T02557; H84734
R.Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence.
A:Reference number: Z14678
A:Accession: T02557
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-175 <ROU>
A:Cross-references: UNIPROT:O80895; EMBL:AC004681; NID:g3298532; PID:g3298546
A:Experimental source: cultivar Columbia
R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H84734
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-175 <STO>
A:Cross-references: GB:AE002093; NID:g3298546; PIDN:AAC25940.1; GSPDB:GN00139
C:Genetics:
A:Gene: T26B15.14; At2g32580
A:Map position: 2
A:Introns: 61/1

Query Match      0.7%; Score 7; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      754 KAERKALV 760
DB      133 KAERKALV 139

RESULT 62
H86805
Hypothetical protein yogM [imported] - Lactococcus lactis subsp. lactis (strain ILL1403)
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: H86805
R.Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: H86805
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-180 <STO>
A:Cross-references: UNIPROT:Q9CFL9; GB:AE005176; PID:g12724438; PIDN:AAK05546.1; GSPDB:G
A:Experimental source: strain ILL1403
C:Genetics:
A:Gene: yogM

```

Query Match 0.7%; Score 7; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 QLLERSI 775
|||||
DB 36 QLLERSI 42

RESULT 63
T32619
hypothetical protein F42A6.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T32619
R:Du, Z.; Scheat, P.; Andrews, S.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of *C. elegans* coemid F42A6.
A:Reference number: Z21201
A:Accession: T32619
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-182 <DUZ>
A:Cross-references: UNIPROT:Q44487; EMBL:AF038613; PIDN:AAB92052.1; GSPDB:GN00022; CESP:
A:Gene: CESP:F42A6.6
A:Map position: 4
A:Introns: 56/1; 106/2

Query Match 0.7%; Score 7; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 QAADGTQ 174
|||||
DB 164 QAADGTQ 170

RESULT 64
A64438
hypothetical protein MJ1106 - *Methanococcus jannaschii*
C:Species: *Methanococcus jannaschii*
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: A64438
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
raon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: A64438
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-185 <BUL>
A:Cross-references: UNIPROT:Q58506; GB:U67553; GB:L77117; NID:gl591744; PIDN:AAB99109.1;
C:Genetics:
A:Map position: REV1046608-1046051

Query Match 0.7%; Score 7; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 536 IYAIIVLP 542
|||||
DB 121 IYAIIVLP 127

RESULT 65
T09390
21K protein precursor - alfalfa
C:Species: *Medicago sativa* (alfalfa)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T09390
R:Coronado, C.
submitted to the EMBL Data Library, February 1997
A:Reference number: Z16658
A:Accession: T09390
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-187 <COR>
A:Cross-references: UNIPROT:P93674; EMBL:Y11553; NID:ei029850; PID:e304664
A:Experimental source: strain A2, subspecies varia; nodule
F:1-18/Domain: signal sequence #status Predicted <SIG>
F:19-187/Product: 21 K protein #status Predicted <MAT>

Query Match 0.7%; Score 7; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 SLSLNKT 127
|||||
DB 63 SLSLNKT 69

RESULT 66
I70176
dynein-like protein 6 - rat (fragment)
C:Species: *Rattus norvegicus* (Norway rat)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 09-Jul-2004
C:Accession: I70176
R:Tanaka, Y.; Zhang, Z.; Hirokawa, N.
J. Cell Sci. 108, 1893-1893, 1995
A:Title: Identification and molecular evolution of new dynein-like protein sequences in
A:Reference number: I55515; MUID:95386588; PMID:7657712
A:Accession: I70176
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-188 <RES>
A:Cross-references: UNIPROT:Q63169; GB:D26497; NID:g871908; PIDN:BAA0505.1; PID:g871909
C:Genetics:
A:Gene: DLP6

Query Match 0.7%; Score 7; DB 2; Length 188;
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 LPDNLKA 282
|||||
DB 132 LPDNLKA 138

RESULT 67
E90187
hA1 protein [imported] - *Sulfolobus solfataricus*
C:Species: *Sulfolobus solfataricus*
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: E90187
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: *Sulfolobus solfataricus* complete genome.
A:Reference number: A99139
A:Accession: E90187
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <KUR>
A:Cross-references: UNIPROT:Q97ZZ0; GB:AE006641; NID:gl3813584; PIDN:AAK40756.1; GSPDB:G
C:Genetics:
A:Gene: SS00432
C:Superfamily: *Methanococcus jannaschii* conserved hypothetical protein MJ0226

Query Match 0.7%; Score 7; DB 2; Length 192;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 SDBLEI 504
|||||

Db 44 SDBLEI 50
|||||

RESULT 68
E75462
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: E75462
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75462
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-195 <WHI>
A:Cross-references: UNIPROT:Q9RV4; GB:AE001942; GB:AE000513; NID:96458611; PIDN:AAF1047
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0897
A:Map position: 1
C:Superfamily: Deinococcus radiodurans hypothetical protein DR0897

Query Match 0.7%; Score 7; DB 2; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 YTSVALL 997
|||||

Db 116 YTSVALL 122
|||||

RESULT 69
F71246
probable indolepyruvate ferredoxin oxidoreductase beta subunit - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 12-Jul-2004
C:Accession: F71246
R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Onfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: F71246
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-196 <KAW>
A:Cross-references: UNIPROT:O57968; GB:AP000001; NID:g3236128; PIDN:BAA29301.1; PID:g325
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0229

Query Match 0.7%; Score 7; DB 1; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 DLSEIKE 506
|||||

Db 115 DLSEIKE 121
|||||

RESULT 70
T29084
3-isopropylmalate dehydratase (EC 4.2.1.33) small chain - Streptomyces coelicolor
C:Species: Streptomyces coelicolor

C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T29084
R:Redenbach, M.; Kieser, H.M.; Denapaita, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopf
Mol. Microbiol. 21, 77-96, 1996
A:Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb
A:Reference number: Z20556; MUID:97000351; PMID:8843436
A:Accession: T29084
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-197 <RED>
A:Cross-references: UNIPROT:O86535; EMBL:AL031124; PIDN:CAA20002.1
C:Genetics:
A:Note: leud
C:Superfamily: 3-isopropylmalate dehydratase small chain
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 0.7%; Score 7; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 FRGSLK 327
|||||

Db 101 FRGSLK 107
|||||

RESULT 71
T44358
hypothetical protein ruva [imported] - Clostridium histolyticum
C:Species: Clostridium histolyticum
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: T44358
R:Matsumura, O.; Jung, C.M.; Katayama, S.; Minami, J.; Takahashi, Y.; Okabe, A.
J. Bacteriol. 181, 923-933, 1999
A:Title: Gene duplication and multiplicity of collagenases in Clostridium histolyticum.
A:Reference number: Z22752; MUID:99121032; PMID:9922257
A:Accession: T44358
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-197 <MAT>
A:Cross-references: UNIPROT:Q9ZNJ6; EMBL:AB014075; NID:g3868863; PIDN:BAA34545.1; PID:g3
A:Experimental source: strain JCM 1403
C:Genetics:
A:Note: ruva
C:Superfamily: holliday junction DNA helicase ruva

Query Match 0.7%; Score 7; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 LELKDKL 523
|||||

Db 126 LELKDKL 132
|||||

RESULT 72
A96504
protein F9C16.19 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96504
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96504
A:Status: preliminary

```
A;Molecule type: DNA
A;Residues: 1-200 <STO>
A;Cross-references: UNIPROT:Q9LNZ9; GB:AB005173; NID:g8778673; PIDN:AAF79681.1; GSPDB:GN
C;Genetics:
A;Gene: F9C16.19
A;Map position: 1

Query Match      0.7%; Score 7; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 QASSSIS 43
Db 188 QASSSIS 194

RESULT 73
I70171
dynein-like protein 1 - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 09-Jul-2004
A;Accession: I70171
R;Tanaka, Y.; Zhang, Z.; Hirokawa, N.
J. Cell Sci. 108, 1883-1893, 1995
A;Title: Identification and molecular evolution of new dynein-like protein sequences in
A;Reference number: I55515; MUID:95386588; PMID:7657712
A;Accession: I70171
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-200 <RES>
A;Cross-references: UNIPROT:Q63164; GB:D26492; NID:g871898; PIDN:BAA05500.1; PID:g871899
C;Genetics:
A;Gene: DUPL

Query Match      0.7%; Score 7; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 276 LPDNLKA 282
Db 144 LPDNLKA 150

RESULT 74
G64006
hypothetical protein HI0370 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
A;Accession: G64006
R;Pleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gockayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: G64006
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-204 <TIGR>
A;Cross-references: UNIPROT:P43989; GB:U32721; GB:L42023; NID:g1573334; PIDN:AAC22028.1;
C;Superfamily: hypothetical protein HI0370

Query Match      0.7%; Score 7; DB 2; Length 204;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 SGSLEQ 117
Db 188 SGSLEQ 194

RESULT 75
```

```
S48952
hypothetical protein YHR110w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004
A;Accession: S48952
R;Latreille, P.
submitted to the EMBL Data Library, May 1994
A;Description: The sequence of S. cerevisiae cosmid 8263.
A;Reference number: S46676
A;Accession: S48952
A;Molecule type: DNA
A;Residues: 1-212 <LAT>
A;Cross-references: UNIPROT:P38819; EMBL:U00059; NID:g529116; PIDN:AAB68853.1; PID:g5291
C;Genetics:
A;Gene: SGD:ERP5
A;Cross-references: SGD:S0001152
A;Map position: 8R
C;Superfamily: conserved hypothetical protein YHR110w
C;Keywords: transmembrane protein

Query Match      0.7%; Score 7; DB 2; Length 212;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 950 SESANSK 956
Db 173 SESANSK 179

RESULT 76
B84206
hypothetical protein vng0475c [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
A;Accession: B84206
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: B84206
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-214 <STO>
A;Cross-references: UNIPROT:Q9HRZ4; GB:AB004437; NID:gi0580082; PIDN:AAG19014.1; GSPDB:G
C;Genetics:
A;Gene: VNG0475C

Query Match      0.7%; Score 7; DB 2; Length 214;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 883 AVATLAD 889
Db 167 AVATLAD 173

RESULT 77
G95069
ABC transporter, ATP-binding protein Vexp2 SP0600 [imported] - Streptococcus pneumoniae
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
A;Accession: G95069
R;Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle,
naon, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: G95069
```

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-215 <KUR>
A;Cross-references: UNIPROT:Q9S4J1; GB:AE005672; PIDN:AAK74752.1; PID:g14972074; GSPDB:G
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0600

Query Match 0.7%; Score 7; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKDK 522
DB 199 TLELKDK 205
|||||

RESULT 78
E97937
hypothetical protein vex2 [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
A;Accession: E97937
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E97937
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-215 <KUR>
A;Cross-references: UNIPROT:Q8DQ3; GB:AE007317; PIDN:AAK99329.1; PID:g15458099; GSPDB:G
A;Gene: vex2

Query Match 0.7%; Score 7; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKDK 522
DB 199 TLELKDK 205
|||||

RESULT 79
C75637
hypothetical protein DRC0023 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
A;Accession: C75637
R;Whice, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: C75637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <WHI>
A;Cross-references: UNIPROT:Q9RZG4; GB:AE001827; NID:G6460959; PIDN:AAF12691.1; PID:G646
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRC0023
A;Map position: plasmid
A;Genome: plasmid
A;Note: plasmid CPI
C;Superfamily: Deinococcus radiodurans hypothetical protein DRC0023

Query Match 0.7%; Score 7; DB 2; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LEEVKLP 476
DB 84 LEEVKLP 90
|||||

RESULT 80
A64479
DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18) - Methanococcus jannaschii
N;Alternate names: endonuclease III
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
A;Accession: A64479
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
rson, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337399; PMID:8688087
A;Accession: A64479
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-220 <BUL>
A;Cross-references: UNIPROT:Q58829; GB:U67584; GB:L77117; NID:g1592077; PIDN:AA899444.1;
C;Genetics:
A;Map position: FOR1403656-1404318
C;Superfamily: Methanococcus jannaschii conserved DNA-(apurinic or apyrimidinic site) lyase
C;Keywords: carbon-oxygen lyase

Query Match 0.7%; Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 869 ILNVEDE 875
DB 67 ILNVEDE 73
|||||

RESULT 81
JQ1986
avrPmaA1 protein - Pseudomonas syringae
C;Species: Pseudomonas syringae
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
A;Accession: JQ1986; S33383
R;Dangl, J.L.; Ritter, C.; Gibbon, M.J.; Mur, L.A.J.; Wood, J.R.; Goss, S.; Mansfield, J
Plant Cell 4, 1359-1369, 1992
A;Title: Functional homologs of the Arabidopsis RPM1 disease resistance gene in bean and
A;Reference number: JQ1985; MUID:93120881; PMID:1477552
A;Accession: JQ1986
A;Molecule type: DNA
A;Residues: 1-220 <DAN>
A;Cross-references: UNIPROT:Q52537; GB:X67808; NID:G45863; PIDN:CAA48009.1; PID:g871512
A;Experimental source: strain P.S.pv maculicola
C;Genetics:
A;Gene: avrPmaA1

Query Match 0.7%; Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ENGEISQ 206
DB 154 ENGEISQ 160
|||||

RESULT 82
JQ1985
avrPp1A1 protein - Pseudomonas syringae
C;Species: Pseudomonas syringae
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004

C;Accession: JQ1985; S33384
 R;Dargl, J.L.; Ritter, C.; Gibbon, M.J.; Mur, L.A.J.; Wood, J.R.; Goss, S.; Mansfield, J.
 Plant Cell 4, 1359-1369, 1992
 A;Title: Functional homologs of the Arabidopsis RPM1 disease resistance gene in bean and
 A;Reference number: JQ1985; MUID:93120881; PMID:1477552
 A;Accession: JQ1985
 A;Molecule type: DNA
 A;Residues: 1-220 <DNA>
 A;Cross-references: UNIPROT:Q52545; GB:X67807; NID:G45864; PIDN:CAA48008.1; PID:G296311
 A;Experimental source: strain P.S.pv pisi race2
 C;Genetics:
 A;Gene: avrPpiA1

Query Match 0.7%; Score 7; DB 2; Length 220;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ENGEISQ 206
 Db 154 ENGEISQ 160
 |||||

RESULT 83

H83639
 hypothetical protein PA0045 [imported] - Pseudomonas aeruginosa (strain PA01)
 C;Species: Pseudomonas aeruginosa
 C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C;Accession: H83639
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A;Reference number: A82950; MUID:20437337; PMID:10984043
 A;Accession: H83639
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-228 <STO>
 A;Cross-references: UNIPROT:Q91787; GB:AE004444; GB:AE004091; NID:G9945863; PIDN:AAG0343
 A;Experimental source: strain PA01
 C;Genetics:
 A;Gene: PA0045

Query Match 0.7%; Score 7; DB 2; Length 228;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 882 LAVATLA 888
 Db 13 LAVATLA 19
 |||||

RESULT 84

C84606
 probable RAS type GTP-binding protein [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
 C;Accession: C84606
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: C84606
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-230 <STO>
 A;Cross-references: UNIPROT:Q9SJ11; GB:AE002093; NID:G4417298; PIDN:AAD20423.1; GSPDB:GN
 C;Genetics:
 A;Gene: At2g21880
 A;Map position: 2
 C;Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match 0.7%; Score 7; DB 2; Length 230;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 490 NNLSKPE 496
 Db 92 NNLSKPE 98
 |||||

RESULT 85

JQ1601
 replication protein C - Clostridium butyricum plasmid pCB101
 C;Species: Clostridium butyricum
 C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
 C;Accession: JQ1601; S18114
 R;Brehm, J.K.; Pennock, A.; Bullman, H.M.S.; Young, M.; Oultram, J.D.; Minton, N.P.
 Plasmid 28, 1-13, 1992
 A;Title: Physical characterization of the replication origin of the cryptic plasmid pCB1
 A;Reference number: PQ0372; MUID:92390513; PMID:1518909
 A;Accession: JQ1601
 A;Molecule type: DNA
 A;Residues: 1-231 <BRE>
 A;Cross-references: UNIPROT:Q45961; EMBL:X62684; NID:G40399; PIDN:CAA44561.1; PID:G40402
 A;Experimental source: strain NCIB7423
 C;Comment: This protein is required for autonomous replication of the plasmid.
 C;Genetics:
 A;Gene: repC
 A;Genome: plasmid
 C;Keywords: plasmid replication

Query Match 0.7%; Score 7; DB 2; Length 231;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 902 KLSQLTS 908
 Db 11 KLSQLTS 17
 |||||

RESULT 86

C72086
 sam-dependent methyltransferase - Chlamydomonas reinhardtii (strain CWL029)
 C;Species: Chlamydomonas reinhardtii, Chlamydia pneumoniae
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
 C;Accession: C72086
 R;Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
 Nature Genet. 21, 385-389, 1999
 A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A;Reference number: A72000; MUID:99206606; PMID:10192388
 A;Accession: C72086
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-238 <ARN>
 A;Cross-references: UNIPROT:Q9Z8G1; GB:AE001622; GB:AE001363; NID:G4376652; PIDN:AAD1852
 A;Experimental source: strain CWL029
 C;Genetics:
 A;Gene: yabc1
 C;Superfamily: methyltransferase, YraI type

Query Match 0.7%; Score 7; DB 2; Length 238;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 595 IPVQAFS 601
 Db 111 IPVQAFS 117
 |||||

RESULT 87

DB6538
 SAM-dependent methyltransferase [imported] - Chlamydomonas reinhardtii (strain J138)
 C;Species: Chlamydomonas reinhardtii, Chlamydia pneumoniae

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C:Accession: D86538
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie
 Nucleic Acids Res. 28, 2311-2314, 2000
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A:Reference number: A86491; PMID:20330349; PMID:10871362
 A:Accession: D86538
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-238 <STO>
 A:Cross-references: UNIPROT:Q5TRU3; GB:BA000008; NID:G8978754; PIDN:BAA98590.1; GSPDB:GN
 A:Experimental source: strain J138
 C:Genetics:
 A:Gene: yabC1
 C:Superfamily: methyltransferase, YxaL type

Query Match 0.7%; Score 7; DB 2; Length 238;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 595 IPVQAFS 601
 |||||
 Db 111 IPVQAFS 117

RESULT 88
 H64343
 hypothetical protein M30352 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
 C:Accession: H64343
 R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
 Reich, C.I.; Overbeck, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1038-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; PMID:96337999; PMID:8688087
 A:Accession: H64343
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-239 <BUL>
 A:Cross-references: UNIPROT:Q57798; GB:U67488; GB:L77117; NID:G2826273; PIDN:AAB98341.1;
 C:Genetics:
 A:Map position: REV3233901-323182

Query Match 0.7%; Score 7; DB 2; Length 239;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 ETLELKD 521
 |||||
 Db 50 ETLELKD 56

RESULT 89
 A86719
 oxidoreductase yhgA [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
 C:Accession: A86719
 R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
 Genome Res. 11, 731-753, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
 A:Reference number: A86625; PMID:21235186; PMID:11337471
 A:Accession: A86719
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-242 <STO>
 A:Cross-references: UNIPROT:Q9CHH8; GB:AE005176; PID:G12723669; PIDN:AAK04851.1; GSPDB:G
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: yhgA

C:Superfamily: NADPH-flavin oxidoreductase homolog

Query Match 0.7%; Score 7; DB 2; Length 242;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 687 DLEKILK 693
 |||||
 Db 29 DLEKILK 35

RESULT 90
 A85688
 hypothetical protein Z1868 [imported] - Escherichia coli (strain O157:H7, substrain EDL)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C:Accession: A85688
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; PMID:21074935; PMID:11206551
 A:Accession: A85688
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-247 <STO>
 A:Cross-references: UNIPROT:O8X3P1; GB:AE005174; NID:G12514793; PIDN:AAG55965.1; GSPDB:G
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z1868
 C:Superfamily: phage P22 gene 18 protein

Query Match 0.7%; Score 7; DB 2; Length 247;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 EVNLPEP 257
 |||||
 Db 96 EVNLPEP 102

RESULT 91
 A84653
 hypothetical protein At2g25810 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 22-Mar-2002
 C:Accession: A84653
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; PMID:20083487; PMID:10617197
 A:Accession: A84653
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-249 <STO>
 A:Cross-references: GB:AE002093; NID:G3643602; PIDN:AAC42249.1; GSPDB:GN00139
 C:Genetics:
 A:Map position: 2
 C:Superfamily: nodulin-26

Query Match 0.7%; Score 7; DB 2; Length 249;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 GNTLVGL 149
 |||||
 Db 47 GNTLVGL 53

RESULT 92

```

AB2991
transcription regulator, GntR family Atu3532 [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AB2991
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2991
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <KUR>
A:CROSS-references: UNIPROT:Q8UA43; GB:AE008689; PIDN:AAL44344.1; PID:gl7741936; GSPDB:C
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3532
A:Map position: linear chromosome

Query Match      0.7%; Score 7; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      755 AKKALVT 761
      |||||
Db      221 AKKALVT 227

RESULT 93
G98292
hypoetical transcription regulator in dcp-nohA intergenic region [imported] - Agrobact
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: G98292
R:Goodner, B.; Hipkile, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: G98292
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <KUR>
A:CROSS-references: UNIPROT:Q8UA43; GB:AE007870; PIDN:AAK89865.1; PID:gi5159808; GSPDB:C
A:Genes: AGR_L_2599
A:Map position: linear chromosome

Query Match      0.7%; Score 7; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      755 AKKALVT 761
      |||||
Db      221 AKKALVT 227

RESULT 94
AB6727
transcription regulator yida [imported] - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: AB6727
R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis se
A:Reference number: AB6625; MUID:21235186; PMID:11337471
A:Accession: AB6727
A>Status: preliminary

```

```

A:Molecule type: DNA
A:Residues: 1-251 <STO>
A:CROSS-references: UNIPROT:Q9CHB5; GB:AE005176; PID:gl2723740; PIDN:AAK04915.1; GSPDB:G
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yida

Query Match      0.7%; Score 7; DB 2; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      591 QLTEIPV 597
      |||||
Db      73 QLTEIPV 79

RESULT 95
A96761
unknown protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96761
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719; PMID:11130712
A:Accession: A96761
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <STO>
A:CROSS-references: UNIPROT:Q9FDZ8; GB:AE005173; NID:gl1120794; PIDN:AAG30974.1; GSPDB:G
C:Genetics:
A:Gene: T9L24.51
A:Map position: 1

Query Match      0.7%; Score 7; DB 2; Length 254;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      747 DLDKXIA 753
      |||||
Db      144 DLDKXIA 150

RESULT 96
WNAD84
early E1A 28K protein - human adenovirus 4
C:Species: Mastadenovirus h4 (human adenovirus 4)
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C:Accession: A25614
R:Tokunaga, O.; Yaegashi, T.; Lowe, J.; Dobbs, L.; Padmanabhan, R.
Virology 155, 418-433, 1986
A:Title: Sequence analysis in the E1 region of adenovirus type 4 DNA.
A:Reference number: A94347; MUID:87071662; PMID:2947381
A:Accession: A25614
A:Molecule type: DNA
A:Residues: 1-257 <TOK>
A:CROSS-references: UNIPROT:P10407; GB:M14918; NID:g209874; PIDN:AAA67091.1; PID:g825435
C:Genetics:
A:Introns: 189/2
C:Superfamily: adenovirus early E1A protein
C:Keywords: DNA binding; early protein; transcription regulation; zinc finger

Query Match      0.7%; Score 7; DB 1; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;

```



```
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 924 QAIQNAA 930
    |||||
Db 129 QAIQNAA 135

RESULT 97
AHI865
hypothetical protein all0473 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AHI865
R;Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AHI865
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <KUR>
A:Cross-references: UNIPROT:Q8Y2I7; GB:BA000019; PIDN:BAE72431.1; PID:gl7129818; GSPDB:G
C:Genetics:
A:Gene: all0473
C:Superfamily: gufa protein

Query Match 0.7%; Score 7; DB 2; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 801 LITGLVE 807
    |||||
Db 182 LITGLVE 188

RESULT 98
S36431
hypothetical protein - Buchnera aphidicola
C:Species: Buchnera aphidicola
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C:Accession: S36431
R;Munson, M.A.; Baumann, P.
submitted to the EMBL Data Library, December 1992
A:Description: tryptophan operon of Buchnera aphidicola.
A:Reference number: S36426
A:Accession: S36431
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <MUN>
A:Cross-references: UNIPROT:P42395; EMBL:Z19055; NID:g396655; PIDN:CAA79503.1; PID:g3966
C:Superfamily: conserved hypothetical protein H1243

Query Match 0.7%; Score 7; DB 2; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 430 VKRKNKL 436
    |||||
Db 59 VKRKNKL 65

RESULT 99
T36359
hypothetical protein SCE9.40 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T36359
R;Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z21575
```

```
A:Accession: T36359
A>Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-258 <OLI>
A:Cross-references: UNIPROT:Q9X8L2; EMBL:AL049841; PIDN:CAB42786.1; GSPDB:GN00070; SCORE
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCORDB:SCE9.40

Query Match 0.7%; Score 7; DB 2; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 999 LITAIKK 1005
    |||||
Db 16 LITAIKK 22

RESULT 100
T19091
hypothetical protein C08F11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T19091
R;Matthews, L.
submitted to the EMBL Data Library, December 1996
A:Reference number: Z19072
A:Accession: T19091
A>Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-259 <WIL>
A:Cross-references: UNIPROT:O62050; EMBL:Z83216; PIDN:CAB05675.1; GSPDB:GN00022; CESP:CO
A:Experimental source: clone C08F11
C:Genetics:
A:Gene: CESP:C08F11.4
A:Map position: 4
A:Introns: 81/3; 132/3; 202/3
C:Superfamily: Caenorhabditis elegans hypothetical protein C33A12.9b

Query Match 0.7%; Score 7; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 LTLTTVS 16
    |||||
Db 184 LTLTTVS 190

Search completed: August 28, 2005, 11:08:51
Job time : 53 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:29:30 ; Search time 173 Seconds
(without alignments)
2253.494 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 5080

Sequence: 1 MKKHLKTVALLTTSVVTH.....LGYTSVALLSLTAIKKKY 1008

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 110 summaries

Database : A_Geneseq_16Dec04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5080	100.0	1008	5	Abp25813 Streptococ
2	5080	100.0	1008	5	Abp53347 Streptococ
3	5080	100.0	1008	8	Adr83923 S. pyogen
4	3777.5	74.4	1055	4	Aau03612 Group B S
5	3777.5	74.4	1055	6	Abp56257 Serotype
6	3776.5	74.3	1055	5	Abp25812 Streptococ
7	243	4.8	4734	8	Adh62807 Lactobaci
8	241.5	4.8	669	8	AdS22318 Bacterial
9	237.5	4.7	668	8	AdS22319 Bacterial
10	220.5	4.3	3257	4	AbS67502 Drosophil
11	217.5	4.3	1072	5	ABBS4963 Lactococc
12	215.5	4.2	641	8	Adk47748 Streptococ
13	215.5	4.2	641	8	Adr94313 Novel S.
14	214.5	4.2	3225	7	AdJ68448 Human hea
15	214.5	4.2	3259	7	AdS56037 Human Pro
16	214.5	4.2	3259	7	AdS56033 Human Pro
17	213.5	4.2	621	6	Abu02658 S. pneumo
18	213.5	4.2	621	6	Abp81493 Streptococ
19	213.5	4.2	621	8	Adh92258 S. pneumon
20	213	4.2	1799	7	AdH87293 Enterococ
21	213	4.2	2045	6	Abu46539 Protein e
22	213	4.2	2045	8	Adr83918 S. pyogen
23	213	4.2	2059	5	Abp25711 Streptococ
24	211	4.2	1881	5	Abp73809 Candida a
25	209.5	4.1	1254	2	Aar07503 Merozoite

Aaw24575 Merozoite
Adi45155 Rice isop
Abu42253 Protein e
Abm73008 Staphyloc
Abj18935 Pathogen
Aau34338 Staphyloc
Aau37403 Staphyloc
Abj19119 Pathogen
Aag82935 S. epider
Aau42656 Protein e
Abp38314 Staphyloc
Ado4851 S epiderm
Ado4803 Staphyloc
Abu15838 Protein e
Abi62804 Methicill
Abu15887 Protein e
Aau34207 Staphyloc
Aau37018 Staphyloc
Aay19935 B. burgdo
Aay19934 B. burgdo
Aar46605 Malarial
Aaw98149 Bacillus
Adr83955 S. pyogen
Abi62805 Methicill
Abp56875 Staphyloc
Abi62792 MRSA FmtB
Abb63519 Drosophil
Abi53116 Protein s
Adk63040 Disease t
Adm18933 Bacterial
Aau34139 Staphyloc
Aau36796 Staphyloc
Aab18161 Plasmodiu
Aar46608 Plasmodiu
Abi53560 Protein s
Adk64380 Disease t
Adg43855 Bacterial
Abb47314 Listeria
Abu33037 Protein e
Abp40235 Staphyloc
Ado6647 Helicobac
Aau35908 Helicobac
Adh62806 Lactobaci
Aau37320 Staphyloc
Abm71190 Staphyloc
Adh62805 Lactobaci
Aau34339 Staphyloc
Abj19059 Pathogen
Aar72826 Human mit
Aau23996 Human mit
Adn95402 Human BSC
Adk70220 Human oes
Adi72172 Human sof
Adi18045 Human sof
Aau99795 Kinetocho
Ado84842 E faecal
Abb60349 Drosophil
Adf45530 Chicken A
Abu43311 Protein e
Adn22762 Bacterial
Adn19417 Bacterial
Abu42252 Protein e
Abp29526 Streptoco
Abb47751 Listeria
Abu32684 Protein e
Ada89551 Staphyloc
Aau37374 Staphyloc
Aau34320 Staphyloc
Abj19002 Pathogen
Abm71899 Staphyloc
Abu07438 Protein d
Adj66483 CenP-F ki
Abb63571 Drosophil

99 186.5 3.7 1312 5 ABB77984
 100 186.5 3.7 1312 6 ABR53887
 101 186.5 3.7 1312 7 ADK63602
 102 186.5 3.7 1312 8 ADN19297
 103 185.5 3.7 1186 8 ADS44694
 104 185.5 3.7 1788 6 ABU18740
 105 185 3.6 2086 4 AAU34143
 106 185 3.6 3818 8 ADP45525
 107 185 3.6 3899 6 ABR92048
 108 185 3.6 3900 8 ADF45500
 109 185 3.6 3907 6 ABR92047
 110 185 3.6 3908 8 ADF45502

ALIGNMENTS

RESULT 1
 ABP25813
 ID ABP25813 standard; protein; 1008 AA.
 XX
 AC ABP25813;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 802.
 XX
 KW Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus pyogenes.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB004789.
 XX
 PR 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;
 PI Tettelin H;
 XX
 DR WPI, 2002-352536/38.
 DR N-PSDB; ABN66444.
 XX

New Streptococcus protein for the treatment or prevention of infection or disease caused by Streptococcus bacteria, such as meningitis, and for detecting a compound that binds to the protein.

Claim 1; Page 3231; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX
 SQ Sequence 1008 AA;
 Query Match 100.0%; Score 5080; DB 5; Length 1008;
 Best Local Similarity 100.0%; Pred. No. 5.6e-301; Indels 0; Gaps 0;
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MKKHLKVALTTLTVSVVTHNQEVFSLVKPILKQTQASSISGADYAESGSKLKINE 60
 Db 1 MKKHLKVALTTLTVSVVTHNQEVFSLVKPILKQTQASSISGADYAESGSKLKINE 60
 Qy 61 TSGPVDVTVDLPSDKRTTPEKIKONLAKPREOELKAVTENTSEKQITSGSOLEQSK 120
 Db 61 TSGPVDVTVDLPSDKRTTPEKIKONLAKPREOELKAVTENTSEKQITSGSOLEQSK 120
 Qy 121 SLSLNTKVPSTSNWEICDFITKGNLTVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVAS 180
 Db 121 SLSLNTKVPSTSNWEICDFITKGNLTVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVAS 180
 Qy 181 FAFTPDKKTAAEYTSRAGENGEISQLDVDGKEIINEGEVFNLSYLLKKVTIPTGYKHIGQ 240
 Db 181 FAFTPDKKTAAEYTSRAGENGEISQLDVDGKEIINEGEVFNLSYLLKKVTIPTGYKHIGQ 240
 Qy 241 DAFVNDKNIABVNLPESLETISDYAFALHALKQIDLPDNLKAIGELAFDNIITGKLSLP 300
 Db 241 DAFVNDKNIABVNLPESLETISDYAFALHALKQIDLPDNLKAIGELAFDNIITGKLSLP 300
 Qy 301 RQLMLABRAFPKSNHIIKTIEFRGNSLVKIGBASQDNDLSQMLPDGLEKIESAFTGNP 360
 Db 301 RQLMLABRAFPKSNHIIKTIEFRGNSLVKIGBASQDNDLSQMLPDGLEKIESAFTGNP 360
 Qy 361 GDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKWLEEDFTYQKNSVT 420
 Db 361 GDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKWLEEDFTYQKNSVT 420
 Qy 421 GFSNKGLOKVRKNQNLBEIPKQHNQVTITEIGDNAFRNVDQNKTLRKVDLEEVKLPSTIR 480
 Db 421 GFSNKGLOKVRKNQNLBEIPKQHNQVTITEIGDNAFRNVDQNKTLRKVDLEEVKLPSTIR 480
 Qy 481 KIGAFAPQSNLKSFEASDDLEETKEGAFNNRLETLELKKVLTIGDAAFIHIIAIV 540
 Db 481 KIGAFAPQSNLKSFEASDDLEETKEGAFNNRLETLELKKVLTIGDAAFIHIIAIV 540
 Qy 541 LPESVQEIGRSAFRQNGANNLIIPMGSKVKTIGEMAFLSNRLHLDLSEKQLTPIVQAF 600
 Db 541 LPESVQEIGRSAFRQNGANNLIIPMGSKVKTIGEMAFLSNRLHLDLSEKQLTPIVQAF 600
 Qy 601 SDNALKEVLLPASLUKTIREEAFKQHLKQLEVASALSIAFNALDDNDGDQFQNKVVVK 660
 Db 601 SDNALKEVLLPASLUKTIREEAFKQHLKQLEVASALSIAFNALDDNDGDQFQNKVVVK 660
 Qy 661 THNSYALADGEHFIIVDPDKLSSTIVLEKILKILEGLDYSTLTQTTQTFQPRDMMTAGKA 720
 Db 661 THNSYALADGEHFIIVDPDKLSSTIVLEKILKILEGLDYSTLTQTTQTFQPRDMMTAGKA 720
 Qy 721 LLSKSNLRQGEKQKFLQEAQPFGLGRVLDLKAIAKAELVTKKATKQQLERSINKAVL 780
 Db 721 LLSKSNLRQGEKQKFLQEAQPFGLGRVLDLKAIAKAELVTKKATKQQLERSINKAVL 780
 Qy 781 AYNNSAIKKNVKKLEKELDLTLTGLVEGKGPLAQATWVGYYLLKPLPEYIYGLNVY 840
 Db 781 AYNNSAIKKNVKKLEKELDLTLTGLVEGKGPLAQATWVGYYLLKPLPEYIYGLNVY 840
 Qy 841 FDKSGKLIYALDMSDTTIGEGQDAYGNPILNVDBDNQGYHALAVATLADYEGLDIKITLN 900
 Db 841 FDKSGKLIYALDMSDTTIGEGQDAYGNPILNVDBDNQGYHALAVATLADYEGLDIKITLN 900
 Qy 901 SKLSQLTSTIRQVPTAAVHRAGIPQAIQNAAEABQLLPKPGTHSEKSSSSSANSKDRGL 960
 Db 901 SKLSQLTSTIRQVPTAAVHRAGIPQAIQNAAEABQLLPKPGTHSEKSSSSSANSKDRGL 960

QY 961 QSNPKTNRGRHSAILPRGSGSGSVYVYLGITVTSVALLSLITAIKKKKY 1008
|||||
Db 961 QSNPKTNRGRHSAILPRGSGSGSVYVYLGITVTSVALLSLITAIKKKKY 1008
|||

RESULT 2

ABP53347
ID ABP53347 standard; protein; 1008 AA.

XX
AC ABP53347;

XX 18-NOV-2002 (first entry)

XX Streptococcus pyogenes BVH-P7 protein SEQ ID NO:2.

DE Streptococcus pyogenes; BVH-P7; antigen; group A Streptococcus; GAS;
KW bacterial pathogen; vaccine; antibacterial; gene therapy; pharyngitis;
KW streptococcal infection; erysipelas; impetigo; scarlet fever; infection;
KW invasive disease; bacteraemia; necrotising fasciitis; toxic shock.
XX
OS Streptococcus pyogenes.

XX Key Location/Qualifiers
FH Peptide 1..21
FT /label= signal
FT Protein 22..1008
FT /label= mature_BVH-P7

XX WO200266650-A2.

XX 29-AUG-2002.

XX 21-FEB-2002; 2002WO-CA000207.

XX 21-FEB-2001; 2001US-0269840P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Martin D, Rioux S, Brodeur BR, Hamel J, Rheault P;

XX WPI; 2002-674948/72.

XX N-FSDB; ABQ81821.

XX New polypeptide useful as a vaccine component for preventing, treating or
diagnosing Streptococcus pyogenes infections, e.g. pharyngitis,
erysipelas, impetigo, scarlet fever, bacteraemia, necrotizing fasciitis or
toxic shock.

XX Claim 17; Fig 2; 52pp; English.

XX The present sequence represents Streptococcus pyogenes BVH-P7 protein
(1). (1) has antibacterial activity and can be used in vaccines and gene
therapy. The Streptococcus pyogenes BVH-P7 polypeptide is useful as a
vaccine component for preventing, treating and/or diagnosing
streptococcal infections, such as pharyngitis, erysipelas, impetigo,
scarlet fever, and invasive diseases such as bacteraemia and necrotising
fasciitis, or toxic shock. A composition comprising the BVH-P7
polypeptide is useful in the manufacture of a medicament for the
prophylactic or therapeutic treatment of streptococcal infection. The BVH
-P7 polynucleotide may be used in designing probes for the detection of
Streptococcus in biological samples. The BVH-P7 polypeptide may also be
used as immunogens for the production of antibodies against streptococcal
infections

XX Sequence 1008 AA;

Query Match 100.0%; Score 5080; DB 5; Length 1008;
Best Local Similarity 100.0%; Pred. No. 5.6e-301;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGHLKTVALLTLTVSVVTHNQVPSLVKPEIPKOTQSSSISGADYAESGSKLKINE 60
|||||
Db 1 MKGHLKTVALLTLTVSVVTHNQVPSLVKPEIPKOTQSSSISGADYAESGSKLKINE 60
|||||

QY 61 TSGPVDDTVTDLFSDKRTTPEKIKONLAKGPREQELKAVTENTESEKQITSGSQLEQSK 120
|||||
Db 61 TSGPVDDTVTDLFSDKRTTPEKIKONLAKGPREQELKAVTENTESEKQITSGSQLEQSK 120
|||||
QY 121 SLSLNKTVPSTSNWEICDFITKGNLTVLGSLSGVEKLSQTDHLVLPQAAADGTQIIOVAS 180
|||||
Db 121 SLSLNKTVPSTSNWEICDFITKGNLTVLGSLSGVEKLSQTDHLVLPQAAADGTQIIOVAS 180
|||||
QY 181 FAFTPDKKTAAEYTSRAGENGESQLDVGDKGIIINEGEVFNLSYLLKKVTIPTGVKHLGQ 240
|||||
Db 181 FAFTPDKKTAAEYTSRAGENGESQLDVGDKGIIINEGEVFNLSYLLKKVTIPTGVKHLGQ 240
|||||
QY 241 DAFVNDKNIAEVLNPESLETISDYAFALHAKQLDLPNLKAIGLAFDFFNQITGKLSLP 300
|||||
Db 241 DAFVNDKNIAEVLNPESLETISDYAFALHAKQLDLPNLKAIGLAFDFFNQITGKLSLP 300
|||||
QY 301 RQLMRLAERAFKSNHIIKTIBFRGNSLVIGASQFQNDLSQMLPDGLEKIESEAFNGP 360
|||||
Db 301 RQLMRLAERAFKSNHIIKTIBFRGNSLVIGASQFQNDLSQMLPDGLEKIESEAFNGP 360
|||||
QY 361 GDDHYNNRVLTWTKSGKNPSGLATENTYVNDKSLWQESPEIDYTKLLEEDFTYKNSVT 420
|||||
Db 361 GDDHYNNRVLTWTKSGKNPSGLATENTYVNDKSLWQESPEIDYTKLLEEDFTYKNSVT 420
|||||
QY 421 GFSNKGKLVKRNKNLEIPKQHGVTITEIGDNAFRNVDFQNKTLRKYDLLEVKLPSTIR 480
|||||
Db 421 GFSNKGKLVKRNKNLEIPKQHGVTITEIGDNAFRNVDFQNKTLRKYDLLEVKLPSTIR 480
|||||
QY 481 KIGAPAFOSNNLKSFEASDDLEETKEGAFMNNRIETLEKDKLVTTIGDAAFHINHIYIV 540
|||||
Db 481 KIGAPAFOSNNLKSFEASDDLEETKEGAFMNNRIETLEKDKLVTTIGDAAFHINHIYIV 540
|||||
QY 541 LPESVQEIERSAFRONGANNLI FMGSKVKTIGEWAFLSNRLEHLDLSQKOLTEIPVOAF 600
|||||
Db 541 LPESVQEIERSAFRONGANNLI FMGSKVKTIGEWAFLSNRLEHLDLSQKOLTEIPVOAF 600
|||||
QY 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSALSHIAFNALDDNDGDEQFNKVVVK 660
|||||
Db 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSALSHIAFNALDDNDGDEQFNKVVVK 660
|||||
QY 661 THNSYALADGEHPIVDPDKLSSTIVLEKILKIEGLDYSTLRTOTTQFRDMMTAGKA 720
|||||
Db 661 THNSYALADGEHPIVDPDKLSSTIVLEKILKIEGLDYSTLRTOTTQFRDMMTAGKA 720
|||||
QY 721 LLSKSNLRQGEKQFLQEAQFPLGRVDLDKAIKAEKALVTKATKNGOLLERSINKAVL 780
|||||
Db 721 LLSKSNLRQGEKQFLQEAQFPLGRVDLDKAIKAEKALVTKATKNGOLLERSINKAVL 780
|||||
QY 781 AYNSAIKKANVKKRLEKELDLTLGLVEGKPLAQATWVGVLKTLPLPEYIIGLVNY 840
|||||
Db 781 AYNSAIKKANVKKRLEKELDLTLGLVEGKPLAQATWVGVLKTLPLPEYIIGLVNY 840
|||||
QY 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVNDEGCHALAVATLADYEGLDIKTILN 900
|||||
Db 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVNDEGCHALAVATLADYEGLDIKTILN 900
|||||
QY 901 SKLSQLSIRQVPTAAVHRAGIFQAIQWAAAEAEQLLPKPGTHSEKSSSSSANSKDRGL 960
|||||
Db 901 SKLSQLSIRQVPTAAVHRAGIFQAIQWAAAEAEQLLPKPGTHSEKSSSSSANSKDRGL 960
|||||
QY 961 QSNPKTNRGRHSAILPRGSGSGSVYVYLGITVTSVALLSLITAIKKKKY 1008
|||||
Db 961 QSNPKTNRGRHSAILPRGSGSGSVYVYLGITVTSVALLSLITAIKKKKY 1008
|||||

RESULT 3

ADR83923

ID ADR83923 standard; protein; 1008 AA.

XX ADR83923;

XX 02-DEC-2004 (first entry)

XX S. pyogenes hyperimmune system reactive antigen Spy0843.
 XX hyperimmune serum reactive antigen; vaccine; anticaline.
 XX Streptococcus pyogenes.
 XX WO2004078907-A2.
 XX 16-SEP-2004.
 XX 02-MAR-2004; 2004WO-EP002087.
 XX 04-MAR-2003; 2003EP-00450061.
 XX (INTE-) INTERCELL AG.
 XX Meinke A, Nagy E, Winkler B, Gelbmann D;
 XX MPI, 2004-653698/63.
 XX N-PsDB; ADR83773.
 XX New isolated nucleic acid molecules encoding hyperimmune serum-reactive
 XX antigens from Streptococcus pyogenes, useful for diagnosing, preventing
 XX and treating S. pyogenes infections.
 XX Claim 14; SEQ ID NO 191; 145pp; English.
 XX This invention describes a novel nucleic acid molecule encoding a
 XX hyperimmune serum reactive antigen or its fragment from Streptococcus
 XX pyogenes. The nucleic acid molecule or hyperimmune serum-reactive antigen
 XX or its fragment are useful for the manufacture of a pharmaceutical
 XX preparation, especially a vaccine, against S. pyogenes infection. In
 XX addition, the hyperimmune serum reactive antigen or fragment is used for
 XX the isolation and/or purification and/or identification of an interaction
 XX partner of the hyperimmune serum reactive antigen or its fragment, for
 XX the generation of a peptide (e.g. anticalines) binding to the antigen or
 XX fragment, or for the manufacture of a functional nucleic acid selected
 XX from aptamers and spiegelmers. The nucleic acid molecule may also be used
 XX for the manufacture of functional ribonucleic acids, such as ribozymes,
 XX antisense nucleic acids and siRNA. ADR83733-ADR84189 represent S.
 XX pyogenes hyperimmune serum reactive antigens, fragments and the encoding
 XX polynucleotide described in the invention.
 XX Sequence 1008 AA;

Query Match 100.0%; Score 5080; DB 8; Length 1008;
 Best Local Similarity 100.0%; Pred. No. 5.6e-301;
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKXHLKTVALLTLTVSVVTHNQEVFSLVKEPILKQTQASSISGADYAESGSKLKINE 60
 DB 1 MKXHLKTVALLTLTVSVVTHNQEVFSLVKEPILKQTQASSISGADYAESGSKLKINE 60
 QY 61 TSGPVDVDTDLFSDKRTTPKIKDNLAKGPRELKAVENTSEKQITSGSOLEOSKE 120
 DB 61 TSGPVDVDTDLFSDKRTTPKIKDNLAKGPRELKAVENTSEKQITSGSOLEOSKE 120
 QY 121 SLSLNTKVPSTSNWEICDFTKGNLTVLGSLGSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
 DB 121 SLSLNTKVPSTSNWEICDFTKGNLTVLGSLGSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
 QY 181 FAPTDPKKTATAEYTSRAGENGEISQLDVGKGIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
 DB 181 FAPTDPKKTATAEYTSRAGENGEISQLDVGKGIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
 QY 241 DAFVNDKNIAEVLNPESLETISDYAFALHAKQIDLPDLNKAIGELAFDNOITGKLSLP 300
 DB 241 DAFVNDKNIAEVLNPESLETISDYAFALHAKQIDLPDLNKAIGELAFDNOITGKLSLP 300
 QY 301 RQLMLAERAFKSNHIKTIIEFRGNSLKVIGEASFQNDLSQLMLPDGLEKIESAFTGNP 360
 DB 301 RQLMLAERAFKSNHIKTIIEFRGNSLKVIGEASFQNDLSQLMLPDGLEKIESAFTGNP 360

QY 361 GDDHYNRVVLWTYKSGKQPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
 DB 361 GDDHYNRVVLWTYKSGKQPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
 QY 421 GFSNKGLOKVRKNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLLEEVKLPTIR 480
 DB 421 GFSNKGLOKVRKNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLLEEVKLPTIR 480
 QY 481 KIGAFAPQSNLKSFEASDDLEIEKEGAFMNNRIETLELKDVLVTIGDAAAPHNHIIYIV 540
 DB 481 KIGAFAPQSNLKSFEASDDLEIEKEGAFMNNRIETLELKDVLVTIGDAAAPHNHIIYIV 540
 QY 541 LPESVQIEGRSAFRQNGANNLIIPMGSKVTKLGEAFLSNRLEHLDLSEQQLTEIPVQAF 600
 DB 541 LPESVQIEGRSAFRQNGANNLIIPMGSKVTKLGEAFLSNRLEHLDLSEQQLTEIPVQAF 600
 QY 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSIAFNALDDNDGDEQFNKVVK 660
 DB 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSIAFNALDDNDGDEQFNKVVK 660
 QY 661 THNSYALADGHEFIVDPDKLSSITVDLEKILKLEGLDYSTLRQTTQTQFRDMTTAGKA 720
 DB 661 THNSYALADGHEFIVDPDKLSSITVDLEKILKLEGLDYSTLRQTTQTQFRDMTTAGKA 720
 QY 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780
 DB 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780
 QY 781 AYNSALTKKANVRKLEKELDLTLGLVEGKGPLAQATWVGYYLLKTPLPPEYIYGLNVY 840
 DB 781 AYNSALTKKANVRKLEKELDLTLGLVEGKGPLAQATWVGYYLLKTPLPPEYIYGLNVY 840
 QY 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDEDNESGYHALAVATLADYEGLDIKTLN 900
 DB 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDEDNESGYHALAVATLADYEGLDIKTLN 900
 QY 901 SKLSQLTSIRQVPTAAVHRAGIFQAIQNAAAEABQLLPKPGTTHSEKSSSESANSKORGL 960
 DB 901 SKLSQLTSIRQVPTAAVHRAGIFQAIQNAAAEABQLLPKPGTTHSEKSSSESANSKORGL 960
 QY 961 QSNPKTNRGRHSALPRTGSGSPVYGIKVTSLVALLSLITAIKKKKY 1008
 DB 961 QSNPKTNRGRHSALPRTGSGSPVYGIKVTSLVALLSLITAIKKKKY 1008

RESULT 4
 AAU03612
 ID AAU03612 standard; protein; 1055 AA.
 XX AAU03612;
 DT 12-SEP-2001 (first entry)
 XX Group B Streptococcus antigenic protein, ID-87.
 DE Group B Streptococcus; encapsulated bacterium; therapeutic; sepsis;
 KW meningitis; neonate; antigenic; vaccine; infection; genital tract;
 KW capsid polysaccharide vaccination.
 XX Streptococcus agalactiae.
 XX WO200132882-A2.
 FN 10-MAY-2001.
 PD 07-SEP-2000; 2000WO-GB003437.
 XX 07-SEP-1999; 99GB-00021125.
 PR (MICR-) MICROBIAL TECHNIQS LTD.
 XX Le Page RWF, Wells JM, Hanniffy SB;

Claim 18; Fig 2; 60pp; English.

PS The present sequence represents a BVH-A4 protein from serotype III Group
XX B streptococcus (GBS) strain COH1 (designated GBS-BVH-A4) (I). (I) has
CC antibiotic and immunostimulant activities, and can be used in polypeptide
CC therapy and in vaccine production. (I) can be used for the therapeutic or
CC prophylactic treatment of GBS bacterial infection in a host susceptible
CC to GBS infection. In particular, (I) is useful for treating or preventing
CC GBS infection in a neonate or infant (e.g. sepsis, meningitis, pneumonia,
CC cellulitis, osteomyelitis, septic arthritis, endocarditis or
CC epiglottitis), in a pregnant woman (e.g. mild urinary tract infection to
CC life-threatening sepsis and meningitis, osteomyelitis, endocarditis,
CC amniotia, endometritis, wound infection (post-caesarean or post-
CC episiotomy), cellulitis or fasciitis), in a non-pregnant adult (e.g.
CC bacteraemia, skin or soft tissue infection, pneumonia, urosepsis,
CC endocarditis, peritonitis, meningitis or emphysema), or in a member of
CC dairy herd (e.g. mastitis). A composition comprising (I) or (I) can also
CC be useful for treating or preventing streptococcal infection

SQ Sequence 1055 AA;

Query Match 74.4%; Score 3777.5; DB 6; Length 1055;

Best Local Similarity 71.7%; Pred. No. 1.7e-221;

Matches 756; Conservative 105; Mismatches 140; Indels 53; Gaps 5;

QY 2 KKHAKTVALTTLTVSVTHNQEVFLVKEPIILKQTOASSISGADYAESGSKSLKINET 61
DB 3 KKHAKTVALTTLTVSVTHNQEVFLVKEPIILKQTOASSISGADYAESGSKSLKINET 61
QY 62 SQPVDVTDVDFSDKRTTPEKIKDNLAKGPRQELKAVTENT-ESEKQITSGSQLEQSK 120
DB 62 NSTVDETVDLFDSDGNSNNSKTESVSDPKQVPKAKPEVTQEAASNSNDASKVEVPQ 121
QY 121 SUSLNTVPSTNSWEICDFTIKGNTLVLSKSGVEKLSQTDHLVLPQAAOGTQLIQVAS 180
DB 122 DTASKKETLETSTWEAKDFTVTRGDTLVGFSKSGINKLSQTSKSHLVLPQAAOGTQL 181
QY 181 FAPTPDKKTAIAEYTSRAGEEISQLDVGDGKEIINEGEVENSLLKVKVTPTGYKHGQ 240
DB 182 FAPTPDKKTAIAEYTSRAGEEISQLDVGDGKEIINEGEVENSLLKVKVTPTGYKHGQ 241
QY 241 DAFVFNKNIAYNLPESETISDYAFALHAKQIDLPNLKAIAGELAFDQNOITGLSLP 300
DB 242 DAFVFNKNIAYNLPESETISDYAFALHAKQIDLPNLKAIAGELAFDQNOITGLSLP 301
QY 301 ROLMLAERAFKSNHKTIEFRGNSLKVIGEASPDNDLSQMLPDGLEKIESEAFPTGNP 360
DB 302 RHLTKLAERAFKSNRIQTVFPLGSKLVIGEASPDNDLSQMLPDGLEKIESEAFPTGNP 361
QY 361 GDDHYNNRVLVTKSGKPSGLATENTYVNDPKSLWQSPEDIDYTKMLEEDFTYQKNSVT 420
DB 362 GDEHYNNRVLVTKSGKPSGLATENTYVNDPKSLWQSPEDIDYTKMLEEDFTYQKNSVT 421
QY 421 GFSNKGLOKVRKNKLEIPKQHNQVITTEIGNAFRNVDFQNTLTKYDLSEVKLPSTIR 480
DB 422 GFSNKGLOKVRKNKLEIPKQHNQVITTEIGNAFRNVDFQNTLTKYDLSEVKLPSTIR 481
QY 481 KIGAFAPQSNNUKSPASDDELEIEKEGAPMNNRIETLEKOKLVITIGDAAFHINHIYIV 540
DB 482 KIGAFAPQSNNUKSPASDDELEIEKEGAPMNNRIETLEKOKLVITIGDAAFHINHIYIV 541
QY 541 LPESVQETGRSFRONGANNITFMGSKVTKTGEMAFNLNLEHLDLSEKQLTETPVOAF 600
DB 542 LPESVQETGRSFRONGANNITFMGSKVTKTGEMAFNLNLEHLDLSEKQLTETPVOAF 601
QY 601 SDNALKEVLLPASLTITREAEFKKHLKQLEVASALSHIAFNALDNDGDGEFQNKVVVK 660
DB 602 SDNALKEVLLPASLTITREAEFKKHLKQLEVASALSHIAFNALDNDGDGEFQNKVVVK 661
QY 661 THNNSYALADGHEFTVDPKSGSTTVLEKILKLEGLDYSFLRTOTTQTFQDMWTAGKA 720
DB 662 THNNSYALADGHEFTVDPKSGSTTVLEKILKLEGLDYSFLRTOTTQTFQDMWTAGKA 721

QY 721 LLSKSNLRQGEKQKFLQEAQPFGRVLDLKAIAKAEKALVTKKATKNGQLLERSINKAVL 780
DB 722 LLSKSNLRQGEKQKFLQEAQPFGRVLDLKAIAKAEKALVTKKATKNGQLLERSINKAVL 781
QY 781 AYNSAISKANVKELEKELDLITGLVSGKGLAQATWVQGVYLLKTPPLPEYIYGLNVY 840
DB 782 AYNSAISKANVKELEKELDLITGLVSGKGLAQATWVQGVYLLKTPPLPEYIYGLNVY 841
QY 841 PDKSGKLIYALDMSDDTTGEGQKDAYGNPILNVDNEDNGYHALAVATLADYEGLDIKTILN 900
DB 842 PDKSGKLIYALDMSDDTTGEGQKDAYGNPILNVDNEDNGYHALAVATLADYEGLDIKTILN 901
QY 901 SKLSQLSIRQVPTAAVHRAGIFQAIQNAEAEAFQQLPKPOTHSKSSSSANSKDRGL 960
DB 902 SSLDKIKAIRQIPLAKYHRLGIFQAIRNAEAEADRLLPK-----TPKGYLNEVPYRKKQM 957
QY 961 QSNPK-----TN-----RGRHSA 973
DB 958 EKNLKPVDYKTPIFNKALPNEKVDGDRRAAKGHNINAEITNSVAVTPIRSEQQLHKSQSDV 1017
QY 974 ILPTGSGKSVYGIYTSVALLSLITAIKKK 1007
DB 1018 NLPQTSSKNFIYBILGVSLCLLFLVTAGKKGK 1051
RESULT 6
ID ABP25812 standard; protein; 1055 AA.
XX AC ABP25812;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 800.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX OS antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
XX PI Tetelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN66443.
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX PS Claim 1; Page 3230-3231; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 1055 AA;

Query Match 74.3%; Score 3776.5; DB 5; Length 1055;
Best Local Similarity 71.7%; Pred. No. 1.9e-221;
Matches 756; Conservative 105; Mismatches 140; Indels 53; Gaps 5;

QY 2 KGHKLTVALTITVSVVTHNQEVSLVKEPILKOTQASSISGADYAESGSKUKINET 61
DB 3 KXHLKTLALATTVSVVTVYQEVYGLREESVKEQQTQSA-SEDDWFEEDNERKTNVSK 61
QY 62 SGPVDDTVTDLFSDKRTTPEKIKDNLAKGPRQELKAVTENT-ESEKQITSGSOLEQSK 120
DB 62 NSTVDETVDLSDGNSNNSSKTESVSDPKQVPKAPKEVTOEASNSNDASKVEVPKQ 121
QY 121 SLSLAKTVPSTSNWEICDPIITKNTLVGLSKSGVEKLSQTDHLVLPQSAADGTQITQVAS 180
DB 122 DTASKKETLETSTWEAKDPVTRGDTLVGFSGINKLSQTSHLVLPASHAADGTQITQVAS 181
QY 181 FAFDPDKTAIAEYTSRAGENEISQLDVDGKEINEGEVFNYSLLKVKTIPTGVKHQIQ 240
DB 182 FAFDPDKTAIAEYTSRAGENEISQLDVDGKEINEGEVFNYSLLKVKTIPTGVKHQIQ 241
QY 241 DAFVDMKNIAEYNLPESLETISDYAFALHALKQIDLPDNLKALGELAFEDNOITGKLSLP 300
DB 242 DAFVDMKNIAEYNLPESLETISDYAFALHMSLKQVLPDNLKALGELAFEDNOITGKLSLP 301
QY 301 ROLMRLAERAFKSNHKTTEFRGNSLVKIGEASFQNDLSQMLPDGLEKISEAFTGNP 360
DB 302 RHLKLAERAFKSNRIQTVEFLGSKLVIGEASFQNDLRNVLDPGLEKISEAFTGNP 361
QY 361 GDHYNRNVLWTKGKNSGLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVT 420
DB 362 GDHYNRNVLWTKGKNSGLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVT 421
QY 421 GFSGKLGOKVKNKLEIKQHNGVTITIGDNAFNVDQNTLARKYDLVLEKLPSTIR 480
DB 422 GFSGKLGOKVKNKLEIKQHNGVTITIGDNAFNVDQNTLARKYDLVLEKLPSTIR 481
QY 481 KIDAFAPQSNLKSFEASDLEIKGAPMNNRIETLEKDKLVITGDAAPHNHYATV 540
DB 482 KIDAFAPQSNLKSFEASDLEIKGAPMNNRIETLEKDKLVITGDAAPHNHYATV 541
QY 541 LPESVQEIQRSPONGANNLIPMGSKVTIGEMAFSLNRLHDLSEQKQTEIPVQAF 600
DB 542 LPESVQEIQRSPONGANNLIPMGSKVTIGEMAFSLNRLHDLSEQKQTEIPVQAF 601
QY 601 SNALKEVLPLASLKTIREAPKQNLKQLEVASALSHITAFNALDNDGDEQFNDKVVYK 660
DB 602 SNALKEVLPLASLKTIREAPKQNLKQLEVASALSHITAFNALDNDGDEQFNDKVVYK 661
QY 661 THNSVALADGHEFIVDPDKLSTIVLEKILKLEGLDYSTLRTQTQFDMWTAGKA 720
DB 662 THNSHMLADGHEFIVDPDKLSTIVLEKILKLEGLDYSTLRTQTQFDMWTAGKA 721
QY 721 LLSKSNLRQGEKQKFIQEAQFGLGRVLDLKAIAKAFAKALVTTKATKNGOLLERSINKAVL 780
DB 722 LLSKSNLRQGEKQKFIQEAQFGLGRVLDLKAIAKAFAKALVTTKATKNGOLLERSINKAVL 781
QY 781 AYNSAIAKANKVRLKELDLTLGLVEGKPLAQATWQGVYLLKTPPLPBYIYGLNYY 840
DB 782 AYNSAIAKANKVRLKELDLTLGLVEGKPLAQATWQGVYLLKTPPLPBYIYGLNYY 841

QY 841 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGTHALAVATLADYEGLDIKTILN 900
DB 842 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGTHALAVATLADYEGLDIKTILN 901
QY 901 SKLSOLTSTIROVPTAAYHRAGIFQAIQNAAEAEOLLKPGTTHSEKSSSESANSKDRGL 960
DB 902 SLSLDTIKAIQIPLAKYHRLGIFQAIQNAAEAEADRLLPK---TPKGYLNEVPYRKKQV 957
QY 961 QSNPK-----TN-----RGRHSA 973
DB 958 EKNLKPVDYKTPFNKALPNEKVDGDRRAKGHNINATNNSVAVTPIRSEQQLHKSQSDV 1017
QY 974 ILPRTGSKGSFVYIGLYTSVALLSLITAIKKKK 1007
DB 1018 NLPQTSSKNNFYIILGYVSLCLLPLVATAGKKGK 1051

RESULT 7
ADH62807
ID ADH62807 standard; protein; 4734 AA.
XX AC ADH62807;
XX DT 15-APR-2004 (first entry)
XX DE Lactobacillus johnsonii mucin binding protein sequence SeqID25.
XX KW mucin binding; probiotic; glycoprotein; epithelial cell;
XX KW gastrointestinal tract; lung; uterine cervix.
XX OS Lactobacillus johnsonii.
XX PN EP1382970-A1.
XX PD 21-JAN-2004.
XX PF 15-JUL-2002; 2002EP-00015609.
XX PR 15-JUL-2002; 2002EP-00015609.
XX PA (NEST) SOC PROD NESTLE SA.
XX PI Pridmore RD;
XX WPI: 2004-111519/12.
XX N-PSDB; ADH62787.
PT Novel mucin binding polypeptide encoded by novel gene of Lactobacillus
johnsonii, useful for binding bacteria to mucin.
XX Claim 8; SEQ ID NO 25; 225pp; English.

This invention relates to novel mucin binding polypeptides and the DNA
sequences which encode them. In particular, the genes are derived from
Lactobacillus johnsonii. The invention may allow a better understanding
of the mechanism by which probiotics (such as L johnsonii) can bind to
mucins, large glycoproteins present on the surface of a large number of
epithelial cells including those in the gastrointestinal tract, the lung
or the uterine cervix. The present sequence is that of an L johnsonii
mucin binding protein of the invention.

Sequence 4734 AA;

Query Match 4.8%; Score 243; DB 8; Length 4734;
Best Local Similarity 19.0%; Pred. No. 5.1e-05;
Matches 230; Conservative 168; Mismatches 429; Indels 386; Gaps 52;
QY 21 NOEVSLVKEPILKOTQASSISGSA-----DYAESG-----KSKLKINETS- 62
DB 3680 DEEKAALQEVSDAQTAAANTALDNTTAAVTEADNGIKAINSIEVPTKSDAKEQATSD 3739
QY 63 --GPVDDTVTDLFSDKRTTPEK---IKDNL-AKGPREQEL-----KAVTENTES 105

Db 3740 LNSAVDEAKAIDQSDNLTDDEBKQVAKQIDSDAKKAQEAIDTAKTNDVDVKKAIIDGTILA 3799
QY 106 -EKQI-----TSGSQLEBQK-----ESLSANKTVPSTSNWEICDPIFKGNTLVGLS 150
Db 3800 IDKDVANNAIDNAVAGKKAEBISPLTDEEKTALNNEVDEKAN-SAKDAINKATTPEGVT 3858
QY 151 KSGVEKLSQTDHLVLPQ-----AADGTQLIQVASPAF---TPDK 187
Db 3859 EAQSGIKSIDDVNPTFSVAKAEAKKAVAEAAEKNAIDSSNLTDEKALKQAEVSDA 3918
QY 188 KTAIAEYTSRAGEGEIIGOLDVDGKEIINEGEVFNYSYLLKKVT-----IPTGYKHIGQ 240
Db 3919 QTAANTAIIDNAITNAEVEAEADNGVKTINGIEVPTKSTTKEQATNDLNNENEVNAKKAIDQ 3978
QY 241 DAFV--DNKNIAEVLNPSLETISDYAFALHAKQIDLPDNL-KAI--GELAF----- 288
Db 3979 DSNLTDEEQVAKQID-----SADKAQADAINNAKTNDVDVKKAIIDGTLLAIDKDVANA 4032
QY 289 -PDNOLITGKLSLPRQLMLRAERAFKSNHIKTIEPRGNSLVKIGBASFDQDNDLSQLMLPDG 347
Db 4033 AIDNAVAGKLEINSLSNNEEQATVDLIN-----NEADNAKQIAEATTPEE 4080
QY 348 LEKIESEFTGNPGDDHNNRVVLTGSKNPSGLATENTYVNDKSLQWSPESIDYTKW 407
Db 4081 VTRAQEE-----GVKNININIVPTTSPAKDAANAAL-----DQALKNKKDEINNATN 4127
QY 408 L-----EEDFTYQNSVTFGSNKGLOKVKGNKLEIPKQHNGVTITEI-----GDNAFRNV 458
Db 4128 ISSEBKTDLIKQATRAANAIAKONINNATNTSEVETAQVDGEKAIADVTGPGLSDIKKESI 4187
QY 459 DPQNTLTKYDUEEVKLPSTIRKIGAFAPQSNLKSFEASDDEBEIKEGAF-MNNRIETL 517
Db 4188 DLINKALNEKQ-DEIN-----NANLSQDESTEIDIOAKKIATEINEINNA 4233
QY 518 ELKDKLVITIGDAFRHINHIYAIPLVPSQVEIGRSA-----FRONGANNLIPIWGSKV 568
Db 4234 QTNDEAKBAADTG--VKNIENVISIP-SIEDACKNATQAIDDALNSKKEINNASNLITDSE 4290
QY 569 KTLGEMAFLSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIRBEAFKKNHLK 628
Db 4291 KT-----DLI--NQATEIA-----NAAKDAINSATNTAVEAAEYKG--- 4325
QY 629 QLEVASALSHIAFNALDNDGDGEQPDNKKVVKVTHNSVALADGEHFIVDPDKLSSTIVDL 688
Db 4326 -----VADINNIHFTNLDS-----KKAANS-AIEDA----- 4351
QY 689 EKILKLIBGLDYSTLRQTQTQTFQDMTTAGKALLSKSLRQGEKQFLQEAQFGLGRVDL 748
Db 4352 -----LTTKKDEINNASNLSDSEKAKLINQA----- 4377
QY 749 DKAIKAKEKALVTKATKNGQLLE-----RSINKAVLAYNNSA-----IKKANV 792
Db 4378 -TEIANAAKAAI--NNATNSAVTAENKGIEDIANINVPSLAETKQAIDAIDAIQVQKAKN 4435
QY 793 KELEKELDILT---GLVEGKPLAQATWQGVYLLKTPPLPEYVIGLVYVFDKSGKLI 848
Db 4436 SQIEEAKNLSDAEQKNLIDQVNNKIAQ---DAINKLNDPATTN-----EVIITDRDKAI 4486
QY 849 VALD--MSDTIGEGQKDAY-----LTTKKDEINNASNLSDSEKAKLINQA-----GNP 868
Db 4487 DQITNLFIPTLDSVQKDAQEAINSAQETKIDINKADNLTDQMKQNLIDQVDQVADKATK 4546
QY 869 ILNVDEDEGHYHALAVATLADYEGLDIDIKTLNSK----- 902
Db 4547 AINNAQTNDVDKAEIEGLEIDISIKVPSLVBESKDDAIKEINDALKKKKTDEINNAADLDQK 4606
QY 903 -----LSQLTSTRQVPTAYHRAQIFQALQNAABAE-----QLLPKCTHSEKSSS 949
Db 4607 QKDELISQITDI-----ATETKTKVFNATTNAEVDABAEAGIKAEVKIIPARTADNNT 4661
QY 950 SESANSKDRGLQSN---PRTNRGRH-----SAILPRTGSKGSFVYGIILGYTSVALLS 998
Db 4662 --ESESKEQTVITNSVQPKRNAVHKNGTPVNNKATLPTQCKKNLTLTLGAAALLGLAG 4719

QY 999 LITAI-----KKKK 1007
Db 4720 VFSLFGLGDKRKK 4732

RESULT 8

ADS22318
ID ADS22318 standard; protein; 669 AA.

XX AC ADS22318;

XX DT 02-DEC-2004 (first entry)

XX DE Bacterial polypeptide #11351.

XX KW Recombinant DNA construct; transformed plant; improved plant property;
cold tolerance; heat tolerance; drought tolerance; herbicide; oomosis;
pathogen tolerance; pest tolerance; plant disease resistance;
cell cycle pathway modification; plant growth regulator;
homologous recombination; seed oil yield; protein yield; carbohydrate;
nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
bacterial polypeptide.

XX OS Bacteria.

XX PN US2003233675-A1.

XX PD 18-DEC-2003.

XX PF 20-FEB-2003; 2003US-00369493.

XX PR 21-FEB-2002; 2002US-0360039P.

XX PA (CAOY/) CAO Y.

PA (HINK/) HINKLE G J.

PA (SLAT/) SLATER S C.

PA (CHEN/) CHEN X.

PA (GOLD/) GOLDMAN B S.

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

WPI, 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide
for expression of a polynucleotide encoding a polypeptide from a
microbial source, useful for producing plants with improved properties.

PS Claim 1; SEQ ID NO 11351; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a
promoter functional in a plant cell, where the promoter is positioned to
provide for expression of a polynucleotide encoding a polypeptide from a
microbial source. The invention also relates to a transformed plant
comprising the recombinant DNA construct and a method of producing a
transformed plant having an improved property. The plant is a crop plant
such as maize or soybean. The method of producing a transformed plant
having an improved property comprises transforming a plant with the
recombinant DNA construct and growing the transformed plant, where the
polynucleotide or polypeptide is useful for improving plant properties.
The recombinant DNA construct is useful for producing plants with
improved plant properties, e.g. improved cold, heat or drought tolerance,
tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
increased resistance to plant disease, better growth rate by modification
of the cell cycle pathway with plant growth regulators, increased rate of
homologous recombination, modified seed oil or protein yield and/or
content, improved yield by modification of carbohydrate, nitrogen or
phosphorus use and/or uptake, by modification of photosynthesis or by
providing improved plant growth and development under at least one stress
condition. This sequence represents a bacterial polypeptide used in the
scope of the invention. Note: The sequence data for this patent did not
form part of the printed specification but was obtained in electronic

CC format from USPTO at segdata.uspto.gov/sequence.html.

XX
SQ Sequence 669 AA;

Query Match 4.8%; Score 241.5; DB 8; Length 669;
Best Local Similarity 22.7%; Pred. No. 4.1e-06;
Matches 124; Conservative 84; Mismatches 214; Indels 125

```

136 ICDFITKNTLVGLSGSGVEKLSQTDHLVLPBSQAADGTQLIQVASFAPTFDPKKTAIAYBT 195
177 LCMSTTHGAATVSSQEQTLPXLRLT--LTVGLTTHSDGTS-VTVTGYSG-SPSDSVTIP--- 229
196 SRACENGEISOLDVDGKEIINEGEVFNLSY--LKKVTIPTCYKXIHGQDAFVDKNNKIAEVN 253
230 -----GEI|||||-----DGLPVTITIGASACAKALTSLTSTVTPIDSVTTIDASAFYGCNTNLISVT 278
254 LPSELETISDYAF-AHLALAKQIDLPDNLKAIGELAFFDQNIQTGKLS--LPRLQMLRAERA 310
279 MPDSVTTIGSSAFYCGCTGLASVTIPNSVTITIGSSAFYG--CTGLISVTIPNSVTITIGSSA 336
311 FKSG-NHUKTIEFRGNSLKVIGE-----ASFQDNDLSQLMLPDGLEKIES 353
337 FYGCTNLTVMMFNENAPTTVGSNWASGTNLVAYVISEGATNFTPEWNEVPYCYPALTAATA 396
354 E-----AFTG-|||:|||:|||:|||-----NPGDDHYNREVLTWTKSGKNPS--GLATENTY--- 388
397 EFTSNIVYGSAPLTVKFTYTGVGANALDWYFDNDSTV-DSTARNPSTVYTPPGTYSIKLN 455
389 -VNPDKLSQESPEID--YTKWLE--BDFTYQKN-----SVTGFSNKGKLVQKRNKNLEIP 439
456 VSNP-----WGSDSEVKYAITVAEPDNFYLSDGTSVITGYSGPG-----GDVVIP 504
440 KQHGVTITIGDNAFNRVDPQNKLRKYDLEEVKLPSTIRKIGAFAPQS--NNLKSPFAS 498
505 STIGNLPVTAISDSVPKS-----NTNITSVTIPDSVTTIGSSAFYCGTGLTSVTIP 555
499 DLLEEIKEGAFMN-NRIETLELKDCLVTIGDAAPH-INHIYAVLPSPVOEIGESAFQON 556
556 NSVTIIGNNAPNGCTGTGTSMTIGNNVTKI DDRPFYGC SALISVWIPDSVTITIGDSAFYQ- 614
557 GANNLIPMGSKVKTGLGEMAFLSNRLHDLSEQKQTEIPVQAFSD--NALKEVLLLPASLK 615
615 -CSNL-----ASVTIPDSVTTIDSYAFRCNALTSTVTPIDSVT 651
616 TIRBEAF 622
652 NVBSRFL 658

```

RESULT 9
ADS22319
ID ADS22319 standard; protein: 668 AA.

AC ADS22319;

DT 02-DEC-2004 ! (first entry)

DE Bacterial polypeptide #11352.

KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.

OS Bacteria.

AA
PN
US2003233675-A1.18-DEC-2003.
PD
XX

XX

20-FEB-2003; 2003US-00369493.

21-FEB-2002; 2002US-0360039P.

(CAO/) CAO Y.
(HINK/) HINKLE G J.
(SLAT/) SLATER S C.
(CHEN/) CHEN X.
(GOLD/) GOLDMAN B S.

Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

WPI; 2004-061375/06.

New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

Claim 1; SEQ ID NO 11352; 122pp; English.

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polypeptide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence, 668 AA;

Query Match 4.7%; Score 237.5; DB 8; Length 668;
Best Local Similarity 20.0%; Pred. No. 7.1e-06;
Matches 128; Conservative 94; Mismatches 207; Indels 211; Gaps 20;

Qy 13 TTTSVVTHNQEVFLSKEPILKQTA\$-----SSISGADYAESGSKLKINETSGPVD 67

```

Qy 68 TVTDLFSDKRTTPEKIKDNLAKGPQELKAVTENTESEKQITSGSQLESLSLNT 127
      :| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 207 SV--FSDC-----SALTAT-----DNDNNSMVASINDGVYVK- 237

```

Qy 128 VPSTSNWEICDFITKGNLTVLGLSKSGVEKLSTQDHLVLPQAADGTOLIQVASFAFTPDK 187
| | | |
Db 238 -----PITTLTLCPS----- 247

QY 188 KTAIAEVTSRAGENGEISQLDVGKEIINEGEVFNSSYLKKVTTIPTGYKHIGQDAFVDNK 247
 : : : : : : : : : : : : : : : :
Db 248 -----GKTGSVT--IPDSVTTIGDSAIFYGCTALTSTVIPDSVTTIGTSAFOGCI 294

QY 248 NIAEVLPESETISDYAFAH-LALKOIDLPDNLKAIGELAFFDNQITCKLSLPRQLMRL-306
:| :|| : : :|| : : :|| : : :|| : : :|| : : :|| : : :|| : : :|| : : :||
Db 295 SITSVNIGNNVTSGNSAFCDICIALTSVIGNNVISGSAFOGCAALTSVTIPDSVTKI 354

```

QY      307 AERAFKSNHIKTIIEFRGNSLVIGEASF-QDNDLSQLMLPD----- 346
      ||:      |      |:      |:      |:      |:      |:      |:

```

Db 355 GSSAFQGCISITSVNIGSGVTSIGNSAFRCQSALTTVTIPDSVTTILDYAFQDCSRLTTV 414
QY 347 ----GLEXIESAFTCNPGDDHYNRVLWTKSGNPSGLATENTYVNPDKSLWQESPEI 402
Db 415 TIGSGVTNIDSSAFSG-----CSALTAID--VDNDSIYASIDGV 452
QY 403 DYTCKWLEEDFTYQKNSVTFSGNKGILQVKVRKNKLEIPKQHGNGVITETIGDNAFRVVD-- 459
Db 453 LYNK-----DITTLIQCPSG-----KTSVPIPD-----SVTTIGDSAFRCDSGLT 493
QY 460 -----FQNKTLKRYDLBVKLPSTIRKIGAFAPQS--NNLKSFASDDEE 503
Db 494 IVPIGNNVTSIGSSAFYGCY----LTSVPIGNNVTSIGSSAFYGCYTSITVIGNNVTS 549
QY 504 IKEGAPMN-NRIETLEKDKLVITIGDAAPH-----INHIYA 538
Db 550 ISSAFRCQSALTVTIPDSVTTIGDIAPFDCSGLTVTVIGNNVTSIGSSAFYGCYALT 609
QY 539 IVPESVQIGRSAFRCQANNLIPMGSKVKTGEMAFLS 578
Db 610 IAPDSVTSIGSADFRCNTLTSVTIGNNVTSIGSSAFRS 649

RESULT 10

ABB67502
ID ABB67502 standard; protein; 3257 AA.

XX ABB67502;

AC ABB67502;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 29298.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-658660/75.

DR N-P8DB; ABL11605.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.

XX Disclosure; SEQ ID NO 29298; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB27072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 3257 AA;

Query Match 4.3%; Score 220.5; DB 4; Length 3257;
Best Local Similarity 18.9%; Pred. No. 0.00071;
Matches 221; Conservative 185; Mismatches 437; Indels 329; Gaps 51;
QY 2 KXHLKTVALTITVSVVTH-----NQBFSLVKEPI--LKQTQAS-----SSISGADYABS 50
Db 1206 KKEAVVGPDLKDTSTSSNTIIDKKSNEFSFDSAMQPSDLNQLKQESAFTKLSISSPKKIMK 1265
QY 51 SGKSKLKNETSGPVDDTVTLFSDKRTTEPKIKDNLAKGPRE--QELKAVTETSEKQ 108
Db 1266 DQDKDLALSKGDSNFITRTGDSRQTDKQHQENTKHEERDSKLKANIDETKSSSE 1325
QY 109 -----ITSGSQLEQSKESLSLNTKVPSTSNWEICDFITKGNLTVGLSKSGVEKLSQTDHL 163
Db 1326 KDAEPIKSDSQSDAKPRLSKPKSRNKKNE-----KKPND--SIAESDIEGGFQVN-- 1376
QY 164 VLPQAAADGTQIQVASFAPP--DKKTAIAEYTSRAGENGESQLDVGKEIINGEV 220
Db 1377 -----TETVQAT--CSTPSSNKKDMVK--SDETNEPNLSETEIG--RIRKRQQA 1421
QY 221 FNSVLLKKVITPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFALHAKQLDLPNL 280
Db 1422 FH-----IENPKDDLHITPQN--ENOSIAGVNF-----KQVPLPESV 1457
QY 281 KAIGELAFFDNOITGKLSLPRQLMLAERAPKSNHIKTIEPRGNSLKVIGASQDDNLS 340
Db 1458 ES-----DTPI--MKIPTKTYLMCTK-----NKTSL--SASEDDPI- 1490
QY 341 QLMPLDGLKIESBAFTGNPGDDHYNRVLVLTSGKNP----- 379
Db 1491 -VLSPQKL--ITTSKGSNPLDNANN--LETSTQDPKEHFSQDTFTDNSDIIPST 1544
QY 380 -----SGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTFSGSKG 426
Db 1545 KKSQIVFPPTPTKSSDQTKNSFITNRSKPSKRVNSKEAKELDNSFEESQNAASESSASK 1604
QY 427 LQK-----VKENKNLEIPKQHGNGVITETIGDNAFRVVDFOKTKRYDLE 471
Db 1605 VKELRPTTASCRLRVLIKRTPTSSLP-----TNSRKSIFKTPAKSKRLTKILES 1656
QY 472 EVKLPS-----TIRKIGAPA-----FQSNLKSFEASDLEIKEGAF 509
Db 1657 MEKTPSREPSVSLGEVNPDSPPVAASVAVLHSDRDLESNEIPNEEVFDETE--ASAED 1715
QY 510 MNNRIETLEKDKLVITIGDAAPHINHIYAIVLPESVQIGRSAFRCQANNLIPMGSKVK 569
Db 1716 TDNKLKKKE-----DDHELVNDICAAASKNPITDDTKDASNSKSTSDVLQETKDE 1767
QY 570 TLGEMAFLSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIRBEAFKKNHLKQ 629
Db 1768 -----LSNSLINATQGEDTPIKELTEEEVPNN-----KTVEDESKQEIILKD 1809
QY 630 LEVASALSHIAFNALDDNDGD-----EQFDNKNVVVTKHNSYALADGHEHFIVDPDKLS-S 683
Db 1810 LEPDNA-----ALEEDTASTAKAABEMDLYIKEKSNVKS-VLAEPETDVTDDLELAQS 1861
QY 684 TIVLEKILKILIEGLDYST---LRQTTQTQPRDMMTAGKALLSKSNLRQ-----GBKQ 733
Db 1862 PIPNSSETTSVTDDEPSTSSVVKRSLRKREADSSQPEA--AKRKORQDVEKSLTGRKE 1919
QY 734 KFLQEAQPFGLGRVLDKAIKAAEALVTKKATKNGOLLERSINKAVLAVNNNSAIKKAN-- 791
Db 1920 QVKPARRQLAEVE--ERPSLKRSTSEAKSTVGQKYISIIIGNETIMSTTAPIRETNRE 1978
QY 792 -----VKRLKEKELDLITG-----LVEGKGPLQAQTMVQGVYLL 824
Db 1979 AASTSPSARKSAVQEAQKHVETTKHILGPPCKILLHSDSPAEEVKKPMVQ--TLLSSTLSL 2037
QY 825 KTPILPLPYIYGLNVYFDKSGKLIYALDMSDTIGEGQKDAYGNFILNVDENEGHALAV 884
Db 2038 QKPSTL-----DDGSPLKIRKLSLKKSIAD-----NIDGQSFSSSSV 2076
QY 885 ATLADYEGLDIKTILNSKLSQLTS-IRQVPTAAVHRAGIFQAIQNAAAEAQQLLPKPGTH 943

Db 2077 --LNKNTSVVAPRKVNISVSLQSKDTQVETAA-----SSSETPLTKKELK 2122
QY 944 SEKSSSESAN---SKRGLOSNP--KTNRGHSAILPR-----TGSKGSFVYGI 988
Db 2123 TQSKTKPEGNKTKSKKSLVQGPOMKQKSEEAIVGPKILNKYLKSETESSRKTSTV 2182
QY 989 LGYTSVALL-----SLITAIAKKKK 1007
Db 2183 TGRKIQIGLEVLKPKPSRKSEESLVAISRKK 2214

RESULT 11
ID ABB54963
XX ABB54963 standard; protein; 1072 AA.
AC ABB54963;
XX
DT 29-AUG-2003 (revised)
DT 16-MAY-2002 (first entry)
XX
XX Lactococcus lactis protein yqfG.
XX
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
XX Lactococcus lactis; IL1403.
XX
XX FR2807446-A1.
XX
XX 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-00004630.
XX
XX 11-APR-2000; 2000FR-00004630.
XX
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
XX WPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification or Lactococcus
PT lactic and related species.
XX
XX Claim 6; SEQ ID NO 1665; 2504bp; French.
XX
XX The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
CC acid sequence is useful in the detection and/or amplification of nucleic
CC acid sequence, particularly to identify Lactococcus lactis or related
CC species. The proteins of the invention are useful for the biosynthesis or
CC biodegradation of a composition of interest. The invention helps research
CC in lactic bacteria, particularly useful in the production of yogurt and
CC cheese. Note: The sequence data for this patent is based on equivalent
CC patent WO200177334 (published 18-OCT-2001) which is available in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX

XX Sequence 1072 AA;
Query Match 4.3%; Score 217.5; DB 5; Length 1072;
Best Local Similarity 16.9%; Pred. No. 0.00023;
Matches 151; Conservative 159; Mismatches 365; Indels 219; Gaps 29;
QY 149 LSKSGVEKLSOT-----DHLVLPQAADGT-----QLIQVASFAFTPDKKT 189
Db 1 MSOKSIKKTMTVGVLTLATLNPFLNQLTLPVEATTSTKITVDEQKITYNLDKVS 60
QY 190 AIAEYTSRAGENG-BISQLDVDGKEIINEGE-----VFNYSLLKKVTIPTGYKHI 238
Db 61 LTAEVASYAGLAGDIVIPDI-----IVNGQTYAITSIGTVAFNSGIRSVIIGNNVVDI 116

QY 239 QGDADFV-----NKNIAEVNLPESLETISDYAFALHAKLQIDLPNKLKAIGELAFPD 290
Db 117 NTSAQTTTAPDYKKSLTKVLTGAKVQNIKTDAFAGNAISSIEFPNSVLKIATRAPAN 176
QY 291 NOITGKLSLPQMLRAEPKSNHIKTIEPRGNSLKVIGBASPODNDLSQMLPDGLEK 350
Db 177 NNLT-ELSLGSGNITEIMAKAFQSNQITIEFADESLLTVDSAAFGSSVQSLLTIGV-- 233
QY 351 IESEAFNCPGDDHYNRVVLWTKSGK-----NPSGLATENTYVNPDKSLWQES 399
Db 234 -----TLADDVFNKTSPLFGQLSDLPDEIRTISVNSSGL-FDKSWINDSS--QS 281
QY 400 PEIDYTKWLEEDFTYQKNSVTGFGNKGLOKVRKNKLEIPKOHNGVTI-----TEIGDN 453
Db 282 TSVS-----TENADTTVPSSSSNSVEQVDVASS-ESTQDANGASLYPISEASVTDN 331
QY 454 AFRNVDFQNTLRKYDLEEVKLPSTIRKIGAFQSNLKKFEASD-----DLEE 503
Db 332 TLNSISSLDSSISSQSTENSQ--SGASSTAEISYDSSENSLSLSSNQINSNSEKDSNQ 389
QY 504 IKGAFNNRRIETLELKDVLTVIGDAAPHIHVAIVLP---ESVQEIGRGAFRONGANN 560
Db 390 SSLGSSGNSNESEHSNSNINETNNSSEIINILPPSNPTESNSVSDQTSSEASTNSNS 449
QY 561 LIPMGSKVTILGEMAFLSNRLEHLDLSEKQLTEIPVOAFSDNALKEVLLPASLKTREE 620
Db 450 ISLSPSNISSTSDSESATNSDFFSNVAEVANNSLASVNNSSSVLSSTSTADNL-GINQS 508
QY 621 AFKNHLKQLEVASALSHIAFNALDDNDGDEQFNKVVVTHHNSYALADGEHFTVDPDK 680
Db 509 GSDNLTKDSSEISTSGAFLSSNQTSE-----ASTNSNS-----SLSLSPSN 550
QY 681 LSSTIVDLEKTLKLEGLDYSTLRQTTQTQFRDMTTAGKALLSKNSLQGEKQKFLQEAQ 740
Db 551 ISSTSV-----LESTTSSNFSNVAEVANNSLASVNNSSSVLSSTSTADN-----LEINQ 601
QY 741 FFLGRVDLDKAIKAELAVTKATKN-----GQLLERSINKAVLAYNNSAIKKANVK 793
Db 602 FGSDNLTKDSSEISTSGAFLSSNQTSSSEASNSMSSINSPSLSLSLTSSESATNQS-- 659
QY 794 RLEKELDLTLGLVEGKPLAQTWVQGVYLLKTLPLPEYVIGLVNVPYDKSGKLIYALDM 853
Db 660 -----SSEATKV-----DNNNS-----STHS 674
QY 854 SDTIGEGOKDAYGNPILNVEDNEGYHALAVATLADYEGLDIKTILNSKLSQLTSTIRQVP 913
Db 675 SNILNSGNSDSDS---DSDDSDS-----SNLSSSPNLETNQTITSSKPSSEVNNISENP 725
QY 914 TAAVHRAGIFQAIQVAAAEARQLLPKPGTHSEKSSSESANSKDRGLOSNPKTN 967
Db 726 KKV-----SSNSVQENSTDHEMSTNPKSS 750
RESULT 12
ID ADK47748
XX ADK47748 standard; protein; 641 AA.
AC ADK47748;
XX
XX 20-MAY-2004 (first entry)
DT
XX Streptococcus pneumoniae protein, Seq ID No 4263.
XX Streptococcus pneumoniae protein, Seq ID No 4263.
KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS Streptococcus pneumoniae.
XX
XX US6699703-B1.
XX
XX 02-MAR-2004.
XX
XX 26-MAY-2000; 2000US-00583110.
XX
XX

Db 700 NNSAFTALSEBDQLL---SQVKELSMVTELRAQVKQLLENNLAEEAQRRLDYESTAH 756
QY 602 DNALKEVL-----LPASLKTIRREAFKONHLKOLEVASALSIAFNALDNDGDGDFDNK- 656
Db 757 DNLLTEQIHSLSIRAKSKDVKEVL-QNELDDVQLQFSEQSTLIRSL-----QSOLQNK 810
QY 657 -----VVVKTTHNSVALADGSHFIVDPDKLSTIVDLKILKILLEGLDYSTLR 704
Db 811 SEVLGAERVRHSKVBELSQALSQKLEITMDQLL-----LEK-----KRDVETLQ 859
QY 705 QTTOFQFDMITAGKALLSKSNLRQGEKQKFLQEAQFFELGRVLDLKAIAKAEKALVTKKA 764
Db 860 QTIEBKDQV-----EISFSTKMWQLNEEKFSLG-VEI-KTLKEQLNLSRAEE 909
QY 765 TNGQLLERSINKAVLAVNNSAIKKA---NVRLKELDLTLGLVEGKPLAQATWQGV 821
Db 910 AKKEQVEDNEVSSGLKQNYDMSAPQISKEBELQHEFDLLKENEQRKRKLQALINRK 969
QY 822 YLLKTPPLPEYIYGLNVYFDSKGLIVALDMSDT-IGBGOKDAYGNPILNVDEDEGSH 880
Db 970 ELLQVRSLREELANK---DESKK---EIPUSETERGEVE-----DKNKEYS 1013
QY 881 ALAVATLADYEGLDIKTLNLSKLSQTSIR-----QVPTAAVHRAGI 922
Db 1014 EKVTSKQCEIEIYIKQISEKEVELQHIRKDLBEKLAEEQFQALVKQMNTLODKTNQ 1073
QY 923 FOAIQNAABAEQLPKPGTTHSEKSSSES-ANSKDRGLQSNPKTNRGH 971
Db 1074 IDLLQAEISENQAIQKLITNTDASDGSVALVKETVVISPPCTGSSEH 1123

RESULT 15

ADSE5037
ID ADE56037 standard; protein; 3259 AA.
XX
AC ADE56037;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human Protein CAA53052, SEQ ID NO 1876.
XX
KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GENO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; CAA53052.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3259 AA;

Query Match 4.2%; Score 214.5; DB 7; Length 3259;
Best Local Similarity 19.8%; Pred. No. 0.0016;
Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;
QY 20 HNEVPSLVKPEILKQTOASSISGADYAESGSGSKLKI-----NETSPVDVTVDL 72
Db 233 HEDELLQLV-----TQAD-----VETEMQKLRLVQLKLEHEESLVGRAVVDL 277
QY 73 FSDKRTTPEKIKDNLAGPREOELK-----AVTENTESEKQITSGSLOESKESLSLNK 126
Db 278 LQQLTAABQRNQLSLOQLQOMAEHNTLNTVTEREESKILLEKELEVAERKLSFH- 336
QY 127 TVPSTSNWEICDFTKGNLVL-----SKSGVEKLSQTDHLVLPQSAADGTQLIQV 178
Db 337 NLQSEMHLLLEQPEAQQAQAELESRYALSALBKHAEMEKTSHLSLQKTG---QELQS 393
QY 179 ASPAFTPKTKTAIAEYTSRAGENCEI-----SOLDVDGKEIINEGEVFNLSLKVVTPT 233
Db 394 ACDALKDQNSKLLQDKNEQAVQSAQTIQQLLEDQLQKSKSEI-----SQFLNRLPLQQ 445
QY 234 GYKHIGQDAFVDNKNIAEAVNLPESLETISDYAFALHALKQIDLPDLNKAIGELAFFDNQI 293
Db 446 -HETASQTSFPDVN-----EGTQVTEENIASLQKRVVEL-----ENEK 484
QY 294 TGKLSLPRQLML-AERAFKSNHKTIEFRGNSLKVIGEASFQDNDSQLMLPDGLEKIE 352
Db 485 GALLLSSTIEELKAENEKLSQITLLEAQNRT---GEA---DREVSEISIVDIANKRS 537
QY 353 SEAFITGPGDDHYNRVVLTSGKNPGLATENTYNPDKSLWQSESEIYTTKWLIEDF 412
Db 538 SSA-----EESGOD-----VLENTFSQKHKL---SVLLEMEKAEQSEI 573
QY 413 TYQKNSTVTG-----FSNKGKLVKVRKNKLE-----IPKQNGVT 446
Db 574 AFLKLOLQKRAEADHEVLDQKEMQWEGEGIAPIKMKVLEDTGQDFPLMPHEESLP 633
QY 447 ITEIGDNFNRVDFQNTLRKYDLEE--VKLPSTIRKIGAFAPQSNLKSFEASDDLEI 504
Db 634 AVE---KEQASTEHQSRSTSEISLNDAGVELKST-----KQDGDKLSLAVPDIQOC 681
QY 505 KEGAFMNNRIETLEKDKLVITIGDAAPH-INHIIVLVPESVQIGR-----SAFRONGA 558
Db 682 HQDELERLKSQILEL-----ELNFKAQEIYEKNLDEKAKESINLNQIEBFFKQAD 733
QY 559 NN-----LIFWGSKVVTGLGEMAFSLNRLHLDL-----SEOKLTETPVOQAFS 601


```

734 NN$SAFTALSEERDOLL----S0VKELSMVTELRAQVKQLEMNLAEBEORRRLDVESQTAH 790
790
602 DNALKEVL-----LPASLKTTIRERAFKKNHLKQLEVASALSHIAFNALDDNDGDBQFONK- 656
791 DNLITEQIHSLSIEAKSKDVKEIVL-QNELDDVOLQFSEOSTLIRSI-----QSOLQNK 844
657 -----VVVKTHNSYALADGEHFIVDPDKLSSTIVDLKEILKIEGLDYSTLR 704
845 SEVLEGAERVHRHISSEKVEELSQALSQKELEITKMDQLL-----LEK-----KEDVETLQ 893
705 QTTQTQFRDMTTAGKALLSKNSLNROGEKQKFLQBAQFFLGRVLDLDKAIAKALVTKKA 764
894 QTIEEKQOQVT-----EISFSWTEKQVQLNEEKFSLG-VEI-KTLKQNLNLSRAEE 943
765 TKNQOLLERSINKAVLAYNNSAIKKA---NVKRLKEKELDLITGLVGEKGPPLAQATWQGV 821
944 AKTEQVEEDNEVSSGLKQNYDEMSPAGQISKEBELQHEFDLLKKENEQRKKLQAALNRRK 1003
822 YLLKTPPLPEYYIIGLVNYPDKSKLIYALDMST-IGEGOKDAYGNPILNVBDNEGYH 880
1004 ELLQVRSLREELANLKK---DESKK---EIPLSETERGEVEE-----DKENKEYS 1047
881 ALAVATLADVEGLDIKILNLSKLSQLTISR-----QVPTAAYHRAGI 922
1048 EKCVTSKCOEIEIYLKOTISEKEVELOHIRKDLEEKLAABEQFQALVKQMNQTLQDKTNQ 1107
923 FQALQNAAAAEQULLPKPGTHSEKSSSES-ANSKDRGLQSNPKTNRGRH 971
1108 IDLLQABISENQAIIOKLITSNTDASDGD$VALVKETVWISPPCTGSGSEH 1157

```

RESULT 16	
AD56033	AD56033 standard; protein; 3259 AA.
XX	
XX	AD56033
XX	AD56033;
XX	
XX	29-JAN-2004 (first entry)
DT	
XX	
XX	Human Protein CAA53052, SEQ ID NO 1872.
DE	
XX	
XX	Human; pain; neuronal tissue; gene therapy;
KW	spinal segmental nerve injury; chronic constriction injury; CCI;
XX	spared nerve injury; SNI; Chung.
KW	
XX	
OS	Homo sapiens.
XX	
XX	WO2003016475-A2.
PN	
XX	
XX	27-FEB-2003.
PD	
XX	
XX	14-AUG-2002; 2002WO-US025765.
PF	
XX	
XX	14-AUG-2001; 2001US-0312147P.
PR	
XX	01-NOV-2001; 2001US-0346382P.
PR	
XX	26-NOV-2001; 2001US-0333347P.
PR	
XX	
XX	(GEHO) GEN HOSPITAL CORP.
PA	(FARB) BAYER AG.
PA	
XX	
XX	Woolf C, D'urso D, Befort K, Costigan M;
PI	
XX	
XX	WPI; 2003-268312/26.
DR	
DR	GENBANK; CAA53052.
XX	
XX	New composition comprising two or more isolated polypeptides, useful for
PT	preparing a medicament for treating pain in an animal.
PT	
XX	
XX	Claim 1; Page; 1017pp; English.
PS	
XX	
CC	The invention discloses a composition comprising two or more isolated rat
CC	or human polynucleotides or a polynucleotide which represents a fragment,
CC	derivative or allelic variation of the nucleic acid sequence. Also

claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at <http://wipo.int/pub/published/pct> sequences.

Sequence 3259 AA:

Db 734 NNSAFTALSEBDQL---SQVKLSMVTELRQVKQLEMLAEARQRRLDYBSQTAH 790
 Qy 602 DNALKEVL---LPASLKTIREAPKQWHLKQLEVASALSHAFNALDDNDGDEQDNK- 656
 Db 791 DNLLTEQIHSLSIEAKSKDVKIEVL-QNELDDVQLQFSEQSTLIRSL-----QSOLQNK 844
 Qy 657 -----VVVKTHNSVALADGSHFIVDPKLSSTIVDLKILKILTEGLDYSLTR 704
 Db 845 SEVLGEARVRHISKVELSQALSKLEITKMDQL-----LEK-----KRDVELQ 893
 Qy 705 QTTQTFPRDMMTAGKALLSKNLROGEKQKFLQEAQFFLGRVDDLDAKAIKAKALVTKKA 764
 Db 894 QTIIEKDDQVT-----EISFSTKQWQVQLNEEKFSLG-VEI-KTLEKEQLNLSRAEE 943
 Qy 765 TNGQLLERSINKAVLAVNSAIKKA---NVKRLSEKLDLLTGLVEGKPLQAQTMWQGV 821
 Db 944 AKKEQVEDNEVSSGLKQNDYMSAPQISKEELQHEFDLLKKEQORRKLQALINRK 1003
 Qy 822 YLLKTPPLPEYVYIGLVVYFDKSGKLIYALDMSDT-IGEGOKDAYGNPILNVDEDENGYH 880
 Db 1004 ELLQVRSLEBELANK---DESKK---EIPUSETERGEVEE-----DKNKEYS 1047
 Qy 881 ALAVATLADYEGLDIKTILNSKLSQTSIR-----QVPTAAVYHRAGI 922
 Db 1048 EKCVTSKQCEIYILKQTISEKEVELQHIRKDLBEKLAABEQFQALVKQMOTLQDKTNQ 1107
 Qy 923 FOAIQNAABAEQLPKPGTGHSEKSSSES-ANSKDRGLQSNPKNRRGH 971
 Db 1108 IDLLQAEISENQAIQIKLITNTDASDGDSDVALVKETVVISPPCTGSSEH 1157

RESULT 17

ABU02658
 ID ABU02658 standard; protein; 621 AA.
 XX
 AC ABU02658;
 XX
 DT 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain protein from coding region #2237.
 XX
 KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae; type 4 strain.
 XX
 PN WO200277021-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-IB002163.
 XX
 PR 27-MAR-2001; 2001GB-00007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masignani V, Tettelin H, Fraser C;
 XX
 DR WPI; 2003-040579/03.
 DR N-P8DB; ABX07949.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX
 PS Claim 1; SEQ ID NO 4474; 56pp; English.
 PS
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the

CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB56454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIFO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 621 AA;

Query Match 4.2%; Score 213.5; DB 6; Length 621;
 Best Local Similarity 23.2%; Pred. No. 0.00019;
 Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31;
 Qy 6 KTVALTTLTVSVV-----THNOVVF-----SLVKEPILK-----Q 35
 Db 3 KTTLSLTAAVILAAVNPNEPILADTPSEVIVKTVGSIQQNNIKYKVLTVGNIGT 62
 Qy 36 TQASSISGADY-ABSSGK-----SKLKNSTSPVDVTVDFLS-----DK 76
 Db 63 VQVNGVTPVEFEAGQDGKPTIPTKIVGDKVTVTEVASQAFSYYPDETGRIVYVPS 122
 Qy 77 RTTEKIKDNLAKPREQLKAVTENTSEKQITSGSQLEQSK----- 120
 Db 123 ITIPSSIKKIQQKGFHSGKAKTII-----FDKGSQLEKIEDRAFDSELEIELPAS 174
 Qy 121 -----SLSLNK-TVPSTSNWEICDPTTKGNTLVGLSKSGVEKLSQTDHLVLP 168
 Db 175 LEYIGTSAPFSQKLKLTFTSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222
 Qy 169 AAD-GTQLIQVA-----SFA-----FTPKKTAIAEYTSRAGENGESIQ 208
 Db 223 VKTLGSNLFRLTSLKXHVDEEGNESFASVDGVLFSDK-KTQLIYVPSQ--KNDES 279
 Qy 209 VDGEIINEGEVFNYSYLLKKVTIPTGYKHQQDAFVNKNIAFVNLPESLETISDYAP-A 267
 Db 280 KETKELASYSFNKNSY-LKKLELNEGLEKIGTGFAPADAIKLEISLPSNLETIERL 338
 Qy 268 HLALKQIDLDPNLKAIGELAF-----FDNQITGKLSLPRQLMLAEAFKSNH 317
 Db 339 NLELKEILDPNVNFGKHWNGLPKLSLTIGNIN---SLPFFSLGVLDLSKEIHK 395
 Qy 318 -----TIEFRGNS--LKVIGEASFQ-DNDLSQMLPDGLEKIESEAPT 357
 Db 396 NKSTEPSVKKQDTPAIPETVKPVVTVSEHKDKLVKSNLSTNDI-----IVEKVDN 451
 Qy 358 GNPQDDHYNRVVLWTK-----SGKNPSSLATENTVVPDKSLM---QSSPEIDY 404
 Db 452 AKP-KKNSNOGVGVGVKDKGLWYVNSGSGSNATGW-----KDKGLWYVNSG 502
 Qy 405 TKWLEED--FTYQKNS---VTGF-SNKG 427

```

Db      503 TGWVKDKGLWYLYNESGSMATGWVKDKGL 531

RESULT 18
ABP81493
ID      ABP81493 standard; protein; 621 AA.
XX
AC      ABP81493;
XX
XX
DT      04-MAR-2003 (first entry)
DE
DE
DE
XX      Streptococcus pneumoniae polypeptide SEQ ID NO 410.
XX
KW      Streptococcus pneumoniae; infection; otitis media; antibacterial;
KW      diagnosis; gene therapy.
XX
OS      Streptococcus pneumoniae.
XX
XX      WO200283855-A2.
XX
XX      24-OCT-2002.
XX
XX      12-APR-2002; 2002WO-US011524.
XX
XX      16-APR-2001; 2001US-0283948P.
XX      18-APR-2001; 2001US-0284443P.
XX
XX      (AMCY ) AMERICAN CYANAMID CO.
XX
XX      Zagursky RJ, Masi AW, Green BA, Chakravarti DN, Russell DP;
XX      Wooters JL;
XX
XX      WPI; 2003-093010/08.
XX      DR      N-PSDB; AB242341.
XX
XX      New Streptococcus pneumoniae polynucleotides, useful for treating or
XX      PT      preventing S. pneumoniae infections, or non-systemic diseases, e.g.
XX      PT      otitis media, which are induced or exacerbated by S. pneumoniae.
XX
XX      Claim 42; Page 677-680; 1091pp; English.
XX
XX      The invention relates to isolated polynucleotides (AB272147-AB242522) of
XX      CC      a Streptococcus pneumoniae genomic sequence, a fragment or degenerate
XX      CC      variant of the polynucleotide or a nucleic acid sequence 95% identical to
XX      CC      one of the polynucleotides. The S. pneumoniae polynucleotides and encoded
XX      CC      polypeptides (ABP81299-ABP81674) are useful for treating or preventing S.
XX      CC      pneumoniae infections or non-systemic diseases, e.g. otitis media, which
XX      CC      are induced or exacerbated by S. pneumoniae. These are also useful for
XX      CC      detecting S. pneumoniae in a biological sample or diagnosing S.
XX      CC      pneumoniae infection in a subject. The polynucleotides have antibacterial
XX      CC      activity and are useful in gene therapy
XX
XX      Sequence 621 AA;

Query Match      4.2%; Score 213.5; DB 6; Length 621;
Best Local Similarity 23.2%; Pred. No. 0.00019;
Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31;

QY      6 KTVALTTLTVSVV-----THNQEVF-----SLVKEPILK-----Q 35
DB      3 KTIILSLTAAVILAAVFPNEPILADTPSEVIKTKVGSIIQQNNIKYKVLTVENIGT 62
QY      36 TQASSISGADY-ABSSGK-----SKLKINETSPPVDDTVTDLFS-----DK 76
DB      63 VQVNGVTPVEHAGQDGPFTPTKITVGDKVFTEVASQAFSYPPDETGRIVYPSS 122
QY      77 RTTPEIKDNLAKPREQELKAVTENSEKOITSGSQLEQSK-----120.
DB      123 ITIPSSIKKIQQKGFPHGSAKTI-----FDKGSQLEKIEDRAFDFSELEEELPAS 174
QY      121 -----SLSLNK-TVPSTSNWEICDPTTKGNTLVGLSKSGVEKLSQTDHLVLP 168
DB      175 LEYIGTSAFSAFSAFSAFSAFSAFSAFSAFSAFSAFSAFSAFSAFSAFSAFSAF 222

```

```

QY      169 AAD-CTQLIQVA-----SFA-----FTPDKTAIABYTGRAGENGEISQLD 208
DB      223 VKTLGSLNLFRLTTLKHYDVEEGNESFASVDGVLFSD-KTQLIYVPSQ--KNDESYPKTP 279
QY      209 VDGKEIINEGEVFNYSLLKKVTIPTGYKHIGODAFVDMKNIAEAVNLPSLETISDYAF-A 267
DB      280 KETKELASYSFNKNSY-LKCLELNEGLEKIGTFAFADAIKLEIEISLPNSLETIERLAFYG 338
QY      268 HLAKQIDLPNLKAIGELAF-----PDNOITGKLSLPRQLMRLAERAFKSNHIK 317
DB      339 NLELKEILPDNVKNFGRHVNGLPKLSLITGNIN---SLPSFLSGLVDSLKEIHIK 395
QY      318 -----TIEFRGNS--LKVIGESFQ-DNDLSQLMLPDGLKIKISEAFT 357
DB      396 NKSTEPSVKKDTFAIPETVKFVVTSEHIKDVLSNLSNDI-----IVEKVDNIQETDV 451
QY      358 GNPDDHYNNRVVLWTK-----SGKNPSPGLATENTYVNPDKSLW---QESPEIDY 404
DB      452 AKP-KKNSNQGVGVWVKDKGLWYLYNESGSMATGWV-----KDKGLWYLYNESGSM-A 502
QY      405 TKWLEED--FTYQKNS---VTGF-SNKGIL 427
DB      503 TGWVKDKGLWYLYNESGSMATGWVKDKGL 531

RESULT 19
ADM92258
ID      ADM92258 standard; protein; 621 AA.
XX
AC      ADM92258;
XX
XX      03-JUN-2004 (first entry)
XX
XX      S pneumoniae antigenic protein sequence SeqID455.
XX
XX      antibacterial; gene therapy; Streptococcus pneumoniae infection;
XX      KW      antigenic.
XX
XX      Streptococcus pneumoniae.
XX
XX      WO2004020609-A2.
XX      11-MAR-2004.
XX      02-SEP-2003; 2003WO-US027401.
XX      30-AUG-2002; 2002US-0407082P.
XX      (TUFT ) UNIV TUFTS.
XX      Camilli A, Hava DL;
XX      WPI; 2004-239189/22.
XX      N-PSDB; ADM92021.
XX      New Streptococcus pneumoniae nucleic acid molecules, useful for
XX      PT      diagnosing, treating and preventing active infections of Streptococcus
XX      PT      pneumoniae.
XX      Claim 27; SEQ ID NO 455; 123pp; English.
XX
XX      This invention relates to novel isolated Streptococcus pneumoniae nucleic
XX      CC      acid molecules and the antigenic polypeptides encoded by them. The
XX      CC      invention may be useful for the production of compounds with an
XX      CC      antibacterial activity or for gene therapy. The nucleic acid molecules,
XX      CC      compositions and methods disclosed are useful for treating Streptococcus
XX      CC      pneumoniae infection. The present sequence is that of an S pneumoniae
XX      CC      protein of the invention.
XX
XX      Sequence 621 AA;

Query Match      4.2%; Score 213.5; DB 8; Length 621;

```

Best Local Similarity 23.2%, Pred. No. 0.00019;
Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31

QY 6 KTVALTITVTSW-----THNQVF-----SLVKPILK-----Q Q35
DB :||| :|
DB 3 KTVILTAAVILAAYVPNEPIADTFSSVIKETKVGSIIQQNNIKKVLTEGNTGT 62
QY 36 TQASSISGADY-AESSGK-----SKLKINETSFPDDTVDTLFS-----DK 76
DB VQVGNGVTPEFEAGQGPKPTIPTKITGVKVFVTVEVASQAFFYPDETGRIVYPPSS 122
QY 77 RTTPKEIKDNLAKGPRQEELKAVTENYESEKQITSGSQLEBSKE-----120
DB 123 ITPSSIKKIQKGFGHGSKAHTII-----FDKGSQLEKIEDRAFDSLEEIELPAS 174
QY 121 -----SLSLNK-TVPSTGNWEICDFITKGNLTGLVSKSGVEKLQSTDHLVLP SQ 168
DB 175 LEVIGTSAPFSQKLKLTFTSSSKLED-----TSHRAPANLSNLEKULTLPSK 222
QY 169 AAD-GTOLIOVA-----SFA-----FTPDKTAIAEYTSRAGENGEISQID 208
DB :||| :|
DB 223 VKTILGSLNFLTTSLKHVDVEEGNESFASVDGVLFSD-KTQLIYPSQ--KNDES YKTP 279
QY 209 VDGEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVNKNATAEVNLPSLETSIDYAF-A 267
DB :||| :|
DB 280 KETNELASYSFNKNSY-LKKLENEGLEGKIGTFAPADAIKLEESLPNSLETIERLAFYG 338
QY 268 HLAKIQDLDPNLKAI GELAF-----FDNOITGKLSLPQLMRLAERAFKSNHIK 317
DB 339 NLELKEILIPDNVKNFGKRVNNGUPKULSTIGNNN---SLPSPFFLSGVLDLSKEIH IK 395
QY 318 -----TIEPRGNS--LKVIGEASFQ-DNDLSQLMLPDGLEKIESEAPT 357
DB :||| :|
DB 396 NKSTEPSVKXDTFAIPETVKFYVTSHEIKDV LKSNLSTSNDI-----IVEKVDNIKQETDV 451
QY 358 GNPQDDHYNNRVLVLTWK-----SGKNPSGLATENTYVNPDKSLW---QSPEIDY 404
DB :||| :|
DB 452 AKP-KKSNQGVVGWVKDKGLWYLINESGSMATGWV-----KDKGLWYLINESGSM-A 502
QY 405 TKMKLEED-PTYQKNS-----VTGP-SNKGL 427
DB :||| :|
DB 503 TGWVKDKGLWYLINESGSMATGWVKDXGL 531

RESULT 20
ADHB7293
ID ADHB7293 standard; protein; 1799 AA.
XX AC
XX ADHB7293;
DT 22-APR-2004 (first entry)
XX Enterococcus faecalis polypeptide #1773.
DE Enterococcus faecalis infection; transcription regulatory element;
KW antibacterial.
XX Enterococcus faecalis.
OS Enterococcus faecalis.
XX US6617156-B1,
XX
XX 09-SEP-2003.
XX
PF 13-AUG-1998; 98US-00134000.
XX
XX 15-AUG-1997; 97US-0055778P.
XX
XX (DOUC/) DOUCETTE-STAMM L A.
PA (BUSH/) BUSH D.
XX
XX Doucette-Stamm LA, Bush D;
PI WPI; 2003-895394/82.
DR

Db 1231 MAKLEAKADADLAKAYGEGVTDIKNQHKSGDPDVARGLHNKSIDEVAQATKD--AITA 1288
QY 385 ENTYNPKSLWQSPEDIDYKWLBE-----DFTYQKNSVTGNSKGLQKVKNK 434
Db 1289 DTTLTAEAKETQRGVNDKEATKAKELAKADADALDKAY-GDGVTISKQNH-----KSGK 1343
QY 435 NLEIPKQHGVTITIEIGDNAPFNVDQ--NKTLRKYDLEEVKLSTPIRKIGAFAPQSN 491
Db 1344 GLDVKKDEHKALEAVAKVTAIEIADPTLTPVREQQKAEVQ----- 1386
QY 492 LKSPF-ASDDLEIEIKEGAFMNNRIETLBKOKLVLTIGDAAFHINHIAIVLPESVQIEGR 550
Db 1387 -KELELATDKIAEAKDA-----DEADKAYGDGVTATENAHVIGKGI-EARKOLAK 1434
QY 551 SAFRQNGANNLIFM-----GSKVTILGEMAFLSNRLEHLDLSEOKQTEPIPVQAFS 601
Db 1435 KDLAEEAAKTKALITIEDKTLTDQKQELLGVDTTEYAKGIENIDAADKAAGVD---KAYS 1491
QY 602 D-----NALKEVLLPAS--LKTIRESAF-----KKHLLKOLEVAS 634
Db 1492 DGVRIILAQYKEGQNLNDRNNAKEFLLEKADKVTKLINDPTLTHDQKVQDQINKVEQAK 1551
QY 635 ALSHIAFNALDNDGDEQFDNKV--VVKTHNSYALADGEHFIVDPDKLSSTIIVDLKIL 692
Db 1552 L---DAIKSVDDAQTADAINDALGKGIENINNQYOHGDG---VDVRKATAK-GDLEKEA 1603
QY 693 KLIIEGLDYSTRQTTQTFQ-RDMTTA-----GKALLS-K 724
Db 1604 AKVKAL---IAKDPTLTQADKQTAQVAADAAKNTAIAAVDKATTTEGINQELGKGITAIN 1660
QY 725 SNLRGEXKQKPOEAQFFLGRVLDLAKAKAEKALVTK-----KATKNGQL----- 770
Db 1661 KAYRFGEGVKARKEA-----AKADLEKEAAKV-KALITINDPTLTRKADKAKQTEAVAKA 1715
QY 771 -----LERSINKAVLAYNNS-----AIKKANVKELEKELDILLTGLVEGK 809
Db 1716 AIAAVDKATTABGINQELGKGITAINKAYRPGEGVKARKEAKADLEREAQVREAIAND 1775
QY 810 GPLAQTWVQGVYLLKTPPLPEYI-----GLNVYFDKSGKLIYALDMSDITGEG-- 860
Db 1776 PTLTRADKAKQTEAVAKAKAIAAIVADKATTAEAGIN---QELGKGITAINKAYRPGEGVE 1832
QY 861 -----OKDAVGNPILNVDEDNQYHALAVAT 886
Db 1833 AHKEAAKANLEKAKETKALISGRYLSSETKAVQKQVQALAKALQVREAKTVEAVK 1892
QY 887 LADYEGLDIKTILNKSOLTSIRQVPTAAVYHRA--GIFQATQNAABE--AEQLLPKPGT 942
Db 1893 LA--ENLGTVAIRSVAAGLAKDTQATPAALNEAKQAIEALKQAAETLAKITTTDAKLT 1950
QY 943 HSEKSSSSSESAN-----SKDGL-----QSNPKTNRGH 971
Db 1951 EAQKAEQSENVSLALKTAIATVRSQAISVKEAKDKGITATRAAYVVPNKAVAKSSSANH 2010
QY 972 SAILPRTSGKSFVYGILGYTSVALLSLITATKKK 1007
Db 2011 ---LPKSGDANSIVLGLGWSL--LLGMVLYSKKKE 2042

RESULT 22

ADR83918 ID ADR83918 standard; protein; 2045 AA.

XX AC ADR83918;

XX DT 02-DEC-2004 (first entry)

XX DE S. pyogenes hyperimmune system reactive antigen Spy0737.

XX KW hyperimmune serum reactive antigen; vaccine; anticaline.

XX OS Streptococcus pyogenes.

XX

PN WO2004078907-A2.
XX 16-SEP-2004.
XX 02-MAR-2004; 2004WO-EP002087.
XX 04-MAR-2003; 2003EP-00450061.
XX (INTE-) INTERCELL AG.
XX Meinke A, Nagy E, Winkler B, Gelbmann D;
PI WPI; 2004-653698/63.
XX N-PSDB; ADR83768.
XX New isolated nucleic acid molecules encoding hyperimmune serum-reactive
PT antigens from Streptococcus pyogenes, useful for diagnosing, preventing
PT and treating S. pyogenes infections.
XX Claim 13; SEQ ID NO 186; 145pp; English.
XX This invention describes a novel nucleic acid molecule encoding a
CC hyperimmune serum reactive antigen or its fragment from Streptococcus
CC pyogenes. The nucleic acid molecule or hyperimmune serum-reactive antigen
CC or its fragment are useful for the manufacture of a pharmaceutical
CC preparation, especially a vaccine, against S. pyogenes infection. In
CC addition, the hyperimmune serum reactive antigen or fragment is used for
CC the isolation and/or purification and/or identification of an interaction
CC partner of the hyperimmune serum reactive antigen or its fragment, for
CC the generation of a peptide (e.g. anticalines) binding to the antigen or
CC fragment, or for the manufacture of a functional nucleic acid selected
CC from aptamers and spiegelmers. The nucleic acid molecule may also be used
CC for the manufacture of functional ribonucleic acids, such as ribozymes,
CC antisense nucleic acids and siRNA. ADR83733-ADR84189 represent S.
CC pyogenes hyperimmune serum reactive antigens, fragments and the encoding
CC polynucleotide described in the invention.
XX Sequence 2045 AA;

Query Match 4.2%; Score 213; DB 8; Length 2045;

Best Local Similarity 18.9%; Pred. No. 0.0011;

Matches 245; Conservative 193; Mismatches 466; Indels 392; Gaps 59;

QY 29 KEPILKQTSASSISGADYAESS-GSKLKLNETSGVDDTVTLDFDKRTPPKIKDNL 87

Db 822 KERUKSDTKQVKDAKNADAIKKAFFEGKVNIPQAHIP-----GDLNKDK-----EKLLAEL 872

QY 88 AKGPREQEL-----KAVNTETSEKQITSGSQLESKESL-----SLNKTVPSTSNWE 135

Db 873 KQKADDTTEKALDVDTLTTEDEKKEQKVTKAELEKAKTDVKNQTTRELDKVPKLK-A 931

QY 136 ICDPITKNTLVGSKSGVEKLSQT-----DHLVLPQAADGTQLIQVASF 181

Db 932 IEDTHVKGN-LEGVKNKAIEDLKKAHTETVAKINGDDTLDKATKEAQVKEADKALAAGKD 990

QY 182 AFT-----PDK-KTAIAEYTSR---AGENGESQLDDVGKEIINE-GEVFNYSLLKKVTIP 232

Db 991 AITKADDADKSVTAVTEHTPKIAAHKTDGDKKAQVDAANTALDKAAEKEGEINKDATLT 1050

QY 233 T--GVKHIGO-----DAFVD--NKNIAEVN----- 253

Db 1051 TEDRAKQLKEVETALTKAKONVKAAKTADAINDARDKGVATIDAVHKAGDGLGARKSQV 1110

QY 254 --LPESLETISDYAPAHALAKQIDLPDNLKAI-----GE-L 286

Db 1111 AKLEAAKATKDKISADPTLTSKEKEQSKAVDAELKKAIEAVNAADTADKVDALGEGV 1170

QY 287 AFFDNQITGKLSLP-----ROLMLAER---AFKSNHIKTIETFRGNSLKVIGEAE---- 332

Db 1171 TDIKNQHKSGDSIDARREAHGKELDRVAQETKGAIEKDPPTLTTEBAKQVKDVKDAKERG 1230

QY 333 -----SFQDNDSLQMLPDGLEKIESBAFTGNPDGDH---YNNRVVLWTKSGKPNPSGLAT 384

Db 1231 MAKLEAKDADALDKAYGEGVTDIKNQHKSGPVDARRGLHNKSIDVAQATKD--AITA 1288
QY 385 ENTYNPDKSLMOESPEIDYTKWLEE-----DFTYQKNSVTGFSNKGLOKVRNK 434
Db 1289 DTTLTAEKETORGVDKATKAKELAKADADALDKAY-GDGVTSIKNQH-----KSGK 1343
QY 435 NLEIPKQHGVTITEIGNAFNRVDPQ---NKTLRKYDLVEVKLPSTIRKIGAFAPQSN 491
Db 1344 GLDVRKDEHKALEAVAKVTAETAEIADPTLTPEVREQQAEVQ-----1386
QY 492 LKSFEE-ASDDEIEIKEGAPMNNRIETLELKDVLVIGDAAPHINHIYALVLPESVQETGR 550
Db 1387 -KELELATDKIAEKDA-----DEADKAYGQGVTAIENAHVIGKI-EARKDLAK 1434
QY 551 SAFRQNGANNLIFM-----GSKVKTIGEMAFSLNRLEHLDLSEQKQLTEIPVQAFS 601
Db 1435 KDLEAAAKTKALIIEDTKLTDDQKEQLLVGVDTEYAKGIENIDAKDAAGVD---KAYS 1491
QY 602 D-----NALKEVLLPAS---LKTIREEAF-----KGNHLKQLEVAS 634
Db 1492 DGVRDILAQYKEGQNLNDRNNAKEFLLKEADKVTKLINDPTLTHDQKVDQINKVEQAK 1551
QY 635 ALSHTAFNALDNDGDEQDNKV--VVKTHNSYALADGEHFIYDPDKLSSTIVLEKIL 692
Db 1552 L---DAIKSVDDAQTADAINDALGKGIENINNOYOHGDG---VDVRKATAK-GDLEKEA 1603
QY 693 KLIEGLDYSTLRQTQTQF-RDMTTA-----GKALLS-K 724
Db 1604 AKVKAL---IAKDPTLTQADKQTAANDAAKNTAIAVDKATTTEGINQELGKGITAIN 1660
QY 725 SNLRGEEKOKFIOEAFFIGRVDLDKATAKAEKALVTK-----KATKNGQL-----770
Db 1661 KAYRPEGVKKARKEA---AKADLEKEAKV-KALITNDPTLTAKADKQTEAVAKALKA 1715
QY 771 -----LERSINKAVLAYNS-----AIKANVKELEKELDLTLGLVEGK 809
Db 1716 AIAAVDKATTAEGINQELGKGITAINKAYRPEGVKKARKEAAKADLREAAKVREAIAND 1775
QY 810 GPLAQATWQGVYLLKTLPLPEYI-----GLNVYFDKSGKLIYALDMSDTIGEG--860
Db 1776 PTLTKADKAKQTEAVAKALKAIAAVDKATTAEGIN---QELGKGITAINKAYRPEGVGE 1832
QY 861 -----QKDAYGNPILNVDEENEGYHALAVAT 886
Db 1833 AHKEAAKANLEKAVKETKALISGRVYLSETERAVQKQAEQALAKALGQVEAAKTVAVK 1892
QY 887 LADYEGLDIKTLNLSKLSOLTSIRQVPTAAYHRA--GIFQAIQNAABE--ABQLLPKPGT 942
Db 1893 LA--ENLGTVAIRSAYVAGLAKDTPQATAALNEAKQAAIEALKQAAAEETLAKITTDKALT 1950
QY 943 HSBKSSSSSESAN-----SKDRGL-----QSNPKTNRGRH 971
Db 1951 EAQKAEQSNVSLAKTAIATVRSQASTASVEAKDKGITARAAYVPNKAVAKSSANH 2010
QY 972 SAILPRTSGSGSVFYGLGYTSVALLSLITAIKKKK 1007
Db 2011 ---LPKSGDANSIVLGLGWSL-LLGWVLYSKKE 2042

RESULT 23
ID ABP25711 standard; protein; 2059 AA.
XX

AC ABP25711;

XX 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 598.

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus pyogenes.
XX WO200234771-A2.
XX
XX PD 02-MAY-2002.
XX
XX PF 29-OCT-2001; 2001WO-GB004789.
XX
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX
XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
XX PI Tettelin H;
XX
XX WPI; 2002-352536/38.
XX N-PSDB; ABN66342.
XX
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX
XX PS Claim 1; Page 3214; 4525pp; English.
XX
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins
XX
XX SQ Sequence 2059 AA;
Query Match 4.2%; Score 213; DB 5; Length 2059;
Best Local Similarity 18.9%; Pred. NO. 0.0011;
Matches 245; Conservative 193; Mismatches 466; Indels 392; Gaps 59;
QY 29 KEPILKQTOASSSISGADYAES--GSKLKLKINETSGPVDVTDLFSDKRTTPEKIKDNL 87
Db 836 KERLKSDTKQVKDAKNDADAIKKAPEEGKVNTPOAHIP-----GDLNKDK-----EKLAL 886
QY 88 AKGPREOEL-----KAVTENTESEKQITSGSOLESKESL-----SLNKTVPSTSNWE 135
Db 887 KQKADDTTEKADVDVTKLTTEDEKKEQKVTKAELEKAKTDVNTQTRELDKKVPLKK-A 945
QY 136 ICDFITKGTNLVGLSKSGVEKLSQT-----DHLVLPSSQAADGTQLIQVASF 181
Db 946 IEDTHVKN-LEGVKNKAIEDLKKHAHTETVAKINGDDTLDKATKEAQVKEADKALAAGKD 1004
QY 182 AFT-----PDK-KTAIAEYTSR---AGENGESQLDVGKEIINE-GEVFNYSLLKKVTIP 232
Db 1005 AITKADDADKVSTAVTEHTPKIKAAHKTGDLKQAVDANTALDKAAAEKERGEINKDATLT 1064
QY 233 T--GVKHIGO-----DAFVD--NKNIAEVN-----253
Db 1065 TEDKAKQLKEVETALTAKQNVKAAKTADALNDARDKGVATIDAVHKAGQDLGARKSQV 1124
QY 254 --LPESLETISDYAFAPHLALKQIDLPONLKA-----GE-L 286

Db 1125 AKLEAAKATKDKISADPTLTSKEKEQSKAVDAELKKAIEAVNAADTADKVDALGEGV 1184
QY 287 AFFDNIQTKGLSLP-----RQLMRLAER---AFKSNHIKTIETPRGNSLKVGBA---- 332
Db 1185 TDIKNQHKSGSDISARREBAHGKELDRVAQETKGAIEKDPPTLTTEBAKQKVDVDAKERG 1244
QY 333 -----SFQNDLSQLMLPDGLEKISEAPTGNPGDDH---YNNRVVLTWKSGKNPSGLAT 384
Db 1245 MAKLEAKADADALDKAYGEGVTDIKNOHKSGDFVDARRGLHNKSIDVAQATKD--AITA 1302
QY 385 ENTYNPNKSLWQESPEIDYTKWLEE-----DFTYKNSVTGFSNKGLOKVKENK 434
Db 1303 DTLTEAEKETQGNVDKEATKAKELAKADADALDKAY-CDGVTSTKNQH-----KSGK 1357
QY 435 NLEIPKQNGVVTITBEIGNAFRNVDFQ---NKTLYDLERVKLPSTIRKIGAFAPQSN 491
Db 1358 GLDVRKDEHKKALEAVAKRVTAIEADPTLTPEVREQQKAEVQ----- 1400
QY 492 LKSPF-ASDLEIEKEGAFMNNRIETLKLKOKLVITIGDAAPHINHIYAIVLPESVQETGR 550
Db 1401 -KELELATDKTAAKDA-----DEADKAYGDVGTATENAHVIGKGI-EARKDLAK 1448
QY 551 SAFRQNGANNLIFM-----GSKVKTLMGEMAFLSNRLEHLDLSEQKQLTEIPVQAFS 601
Db 1449 KDLEAAAKTKALIETDKLTDDQKEQLGLGVDTYAKGIENIDAKDAAGVD---KAYS 1505
QY 602 D-----NALKVLLPAS---LKTIRSEAF-----KONHLKOLEVAS 634
Db 1506 DGVDRILAQYKEGQNLNDRNAKEFLLEKADKVKTLNDPDLTHDQKVDQINKVEQAK 1565
QY 635 ALSHTAFNALDDNDGQFQDNKV--VVKTHNSYALADGEHFVDPDKLSSTIVDLKIL 692
Db 1566 L---DAIKSVDDAQTADAINALGKGIENINNOYQHGDS---VDVRKATAK-GDLEKEA 1617
QY 693 KLIEGLDYSTLRQTQTQF-RDMTTA-----GKALLS-K 724
Db 1618 AKVKAL---IAKDPLTQADKQKQTAANDAAKNTAIAAVDKATTTEGINQELGKITAIN 1674
QY 725 SNLRGKQKQFQIQAFFLGRVLDLKAIAKAEKALVTK-----KATKNGQL----- 770
Db 1675 KAYRPGEGVKARKEA---AKADLEKEAAKV-KALITNDPTITKADKAKQTEAVAKALKA 1729
QY 771 -----LERSINKAVLANN-----AIKKANVKELEKELDLTLGLVEGK 809
Db 1730 AIAAVDKATTAGINQELGKITAINKAYRPGEGVKARKEAAKADLEKEAAKVRKAIAND 1789
QY 810 GLPAQATVQGVYLLKTLPLPEYI-----GLNVYFDKSGKLIYALDMSDTIGEG-- 860
Db 1790 PTLTKADKAKQTEAVAKALKAIAAIVDKATTAGIN---QELGKITAINKAYRPGEGVE 1846
QY 861 -----QKDAYGNPILNVDEDNEGYHALAVAT 886
Db 1847 AHKKAANLEKAVAKETKALISGDRYLSETEKAVQKQVQALAKALQVQEAATVEAVK 1906
QY 887 LADYEGLDIKTILNSKLQSLTISIRQVPTAAHYRA--GIFQATONAAE--AEQLLPKPGT 942
Db 1907 LA--ENLGTVAIRSAVAGLAKDQDQATAALNEAKQAAIEALQKAAETLAKITTDALIT 1964
QY 943 HSEKSSSESAN-----SKDRGL-----QSNPKTNRGRH 971
Db 1965 EAQKAEQSENVSLALKTAITATVRSQAQSTASVKEAKDKGITAIRAAVVPNKAVAKSSANH 2024
QY 972 SAILPRTSGKGFVYGIIGYTSVALLSLITAKKKK 1007
Db 2025 ---LPKSGDANSIVLVLGLVMSL-LLGWLVYKKKE 2056

RESULT 24
ABP73809
ID ABP73809 standard; protein; 1881 AA.
XX
AC ABP73809;
XX

DT 30-JAN-2003 (first entry)
XX Candida albicans essential protein SEQ ID NO 7646.
DE Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
XX signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
KW Candida albicans.
XX OS
XX WO200253728-A2.
PN 11-JUL-2002.
PD 26-DEC-2001; 2001WO-US049486.
PF 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX (ELIT-) ELITRA PHARM INC.
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
PI WPI; 2002-566694/60.
DR N-PSDB; ABZ32359.
XX Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX Claim 44; SEQ ID NO 7646; 167pp + Sequence Listing; English.
PS The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
SQ Sequence 1881 AA;

Query Match 4.2%; Score 211; DB 5; Length 1881;
Best Local Similarity 19.9%; Pred. No. 0.0012;
Matches 227; Conservative 173; Mismatches 431; Indels 310; Gaps 48;
QY 5 LKTVALTUTTVSVVTHN-----QEVFSLVKEPIKQTOASSSIS 43
Db 689 LPTVLTVDVYVNLVKDNLFLRIKRALFHPDPSPSGPKISYELFEHLDTKVNLTKELQ--T 746
QY 44 GADYAESGSKGKIKINETSFGPVDDTVTDLFSQDKFTTPEKIKDNKLAGPREQELKAVTNT 103
Db 747 EKENAESNDK---ELNEKI-----EKLTNLSTKLET---KLED-----KEQELAKIQSDH 790


```
Db 403 YDAEELNNDKSTAKVLIEINLESVKHNLSEITNIKQGEKIYSKAKDIMQIKATSE 462
QY 359 NPG-----DDHYN-----NRVLWTSGKNPSPGLATENTVYVNDPKSLMOES 399
Db 463 NTAETKLEKVDQSNVYNYLNIQTERRNLIVTEKNRLN-----GIDSTITIEGAL-KES 517
QY 400 P-----BIDYKMLEE-----DFTYQK-NSVTG-----FSNKGLOKVKRKNL----- 436
Db 518 KGNYSIGLEKLEETGKRNKLVDTYKKSINSVTGNSFSLFNNFDLNIQYDFNKNINDYEN 577
QY 437 ---EIPKQHGVTITEIGNAFRNV-----DFQN-KTLR-----KYDL-----EEVKLP 476
Db 578 YNGEYNEPEG-SLNKISEN-LRNASENTSDYNSAKTLLEAQEKVNLNKEEANKYL 635
QY 477 STIRKIGAFAPSNLKSFEASDDLEETKEGAFMNRRIETLELKVLTIVTI-----GDAAPH 532
Db 636 RDVKVSGFRP-----IFNMKESLDKINE-----MIKKEQLTVNEGHGNVKQL 678
QY 533 INHIYAVLPESVQIGRSAPFQNGANNLI-----FMGSKVTLIGEMAFLSNRLEHLDLSEQ 589
Db 679 VENIKELVDENNLSDILQATOKNEEIOKITHSLKNKAKTI-----LGHVDTSAK 729
QY 590 KQLTIPQAFSDNALKEVLLPASLKTIREEAPK-KNHL-----KOLEVASALS 637
Db 730 ----YVGIKITPELALTELLGDAKLKTAGELKFESKNVVLATENMKNNTNLDVHKNIQ 785
QY 638 ---HIAFNALDND-----GDB-----QFDNKV-VVKTHHNSYALA 669
Db 786 DAYVALEILAHSDIEDTKQDSSKLIEMGNQIYLVKVLINGYKNKISSIKSKEAVSVK 845
QY 670 DG-----EHP-----IVDPKLSSTIVDLKILKLEGLDYSTLRQTT-----OTQ 710
Db 846 IGVNSKKHSELKSTICSDSKSYDNIILAEKQTEL-QNLNSFTQEKNTNSDSKLEKIKTD 904
QY 711 PRDMTATAGALLSKSN-----LROGEKQKFLQEAQFGLGRVDL 748
Db 905 FESLNALKTLGEVNALKASDNHEHVQSKSEPNPALSEIEK-----EETDIDSLNTAL 960
QY 749 DKAIAKAE-----KALVTKATKNGQL-----LERSINKAVLAYNNSAIKKANVKRL 795
Db 961 DELLKKGRCTCESRVYKLIKNDVTKEISDDTELTINTIEKNV-KAYLAY-----IKKNYEDTV 1015
QY 796 EKELDLILGLVEGKPLAQATWQGVYLLKTPLPPEYVIGLV-----YFDKSGK--- 846
Db 1016 Q----DVLT-----LNEHFNTKQVSNHBEFTNPDKSNKSSB 1046
QY 847 -LIYALDMSDTIGEGQDAYGNPILNVDEBNEGYPHALAVATLADYEGLDIKTILNSKLSQ 905
Db 1047 ELTKAVTDSKTIISKLKGV-----LIEVNENTE-----MNTIESSAKEIEALYNELONK 1095
QY 906 LRSIRQV 912
Db 1096 KTSLENEI 1102

RESULT 26
AAW24575
ID AAW24575 standard; protein; 1254 AA.
XX
AC AAW24575;
XX
DT 25-MAR-2003 (revised)
DT 10-NOV-1997 (first entry)
XX
DE Merozoite apical-end protein clone 5.3.
XX
KW Merozoite apical-end protein; MAEP; Plasmodium vivax; antigen; malaria;
KW human; erythrocyte; antimalarial vaccine; antibody; blood stage parasite;
KW Duffy blood group antigen; red blood cell; therapy.
XX
OS Plasmodium vivax.
XX
```

```
PN US5646247-A.
XX
PD 08-JUL-1997.
XX
PF 04-OCT-1991; 91US-00792865.
XX
PR 05-APR-1989; 89US-00334041.
PR 06-APR-1989; 89US-00334270.
PR 03-APR-1990; 90MO-US001849.
PR 02-NOV-1990; 90US-00608639.
XX
PA (UTNY ) UNIV NEW YORK STATE.
XX
PI Galinski MR, Barnwell JW;
XX
DR WPI; 1997-362995/33.
XX
N-PSDB; AAT80072.
XX
Plasmodium merozoite apical end protein - useful as antigen for
production of anti-malarial vaccines.
XX
Claim 1; Col 29-38; 68pp; English.
XX
AAW24575 and AAW24576 represent the merozoite apical end proteins (MAEP)
isolated from two different Plasmodium vivax strains. These proteins are
the antigens of the invention, and immunoreact with antibodies against a
native MAEP sequence. P. vivax is one of the four malarial species that
infects humans, and is difficult to target for a vaccine, as it cannot be
cultured in vitro. The preinvasion orientation of malarial merozoites
indicates that the apical end plays an important role in the invasion
process. The MAEP protein binds to the surface of susceptible
erythrocytes from P. vivax susceptible humans and primates, and also
binds to rabbit erythrocytes. The antigen can be used for the production
of antimalarial vaccines. The antigens are involved in the invasion
process, and are immunochemically reactive with antibodies raised against
malaria (particularly P. vivax) blood stage parasites. Synthetic
proteins, polypeptides, peptide fragments and analogues of these antigens
can be used similarly. As the antigens specifically bind to a Duffy blood
group antigen (the antigen present on the surface of susceptible
mammalian red blood cells), and are necessary in the process of invasion
of red blood cells by merozoites, they can be used to inhibit the
invasion of red blood cells by a malarial organism. The antigens can also
be used in a method for inhibiting invasion of susceptible mammalian
blood cells by malarial merozoites, and in a method for inhibiting the
propagation of a malarial organism in susceptible red blood cells.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 1254 AA;
Query Match 4.1%; Score 209.5; DB 2; Length 1254;
Best Local Similarity 21.3%; Pred. No. 0.00086;
Matches 244; Conservative 176; Mismatches 350; Indels 377; Gaps 66;
QY 23 EVFSLVKEPILKQTQASSISGAD-----YAE-----SSGSKLKINETSGPVDDTVDLFS 74
Db 76 EKILYKKEIDEIKQKTNEYKQGTSNFYTYEQYNSATQSKAKIEQ-----FINIAT 127
QY 75 DKRTTPKIKDNLAGPREQELKAVTENTSEKIT---SGSQLEQSKESLSLNTKVPST 131
Db 128 TKKGTSDTSQD-----INELESIKBEVHKNLQLVKQESNMEEMRKQILSKDLLILN 180
QY 132 SNWEICDFITKG-NTLVGLSKSGVEKLSQTDHLVLPQADGTQLIQVASFAFTPKDKTA 190
Db 181 NSETIAKESINNTQNALGFRENAKTKLTKTDEL-LQRVAA-----MIEAK-----AHQNN 230
QY 191 IARYTSRAGENGESQLDVGKEIINEGEVNSYLLK-----KVTIPTGYKHIGQD--A 242
Db 231 IDIALEDAQIDTEYSKIEQINREIMNKKDEIKSYLSEIKYKDKCTTEISNKGKOKIE 290
QY 243 FVD-----NK-NIAEVNLPESLETISDYAFALHAKQI-DLPDLNKAIGEL----- 286
Db 291 FLEKFKPNEESNKNVINEIN--ENIRNSQY-----LKDIEDAKQASTKVELPFKH 342
```



```
CC invention.
XX SQ Sequence 1905 AA;
Query Match 4.1%; Score 207; DB 8; Length 1905;
Best Local Similarity 20.5%; Pred. No. 0.0022;
Matches 208; Conservative 150; Mismatches 366; Indels 290; Gaps 44;
QY 29 KEPILKQTAQSSISGADYAEBSG-----KSKLKINETSPPVDDIVTDLFSDKRTPP 80
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
730 KNDMTMQLESQSQAGDDSHKTRSLLSBAQLHKEBELNLKS-----LNDLHVESKTAA 783
QY 81 EKIKONLAK-GPREQLKAVTEN-----TESKQITSGSQ-----LEQ-----SKESL 125
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
784 ESALQRIAELETQVQELSAEBSLHTEFESKLSAEKSKMSDLEQELKDATECNSR 843
QY 126 KTVPTS-NWEICDPTITGNL-----VGLSKSG-----VEKLSQ 159
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
844 LRVDELSELEBESSLTKGALEHATSKKIDLEALYQSILLEDETMKLOQAGENLTQKETE 903
QY 160 THLVLPQAGDTQLIQVAFPTPKDKTAIAEYTSRAGENGESQLDVQKGIINIGE 219
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 QOELSEKLKAEE-----QAASY-----QAKATAAE-----EVESVKVELEAFETIS 947
QY 220 VNSYLLKKVPTPTGYKHIGQDAFVNDKNIAEVNLPESLETISDYAFALHAKQIDLPDN 279
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
948 TLETTIEELKTKASNAEGRAEQALVESAMMSETN-----QALKE-DLDAK 991
QY 280 LKAIGEL-AFFDNQITGKLSPLQMLRAERAFK--SNHIKTIEF-----RG----- 323
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
992 LAMLELQEQFDSHAEK-----EVEFTKLSAHEKTIEHLTEVHSGLELHATA 1040
QY 324 -----NSLKVIGEASFQNDLSQLMPDGLK-KIHESAFTGNPGDDHNNRVV 370
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1041 ESKNAELEAQMEHALETIGKDSVKDLNERLA--ALESEITSLTHVNEAMKQEINAKVL 1098
QY 371 LMTGKQKPSGLATENTYVNPDKSLWQSPK-----IDYTKWLEEDFTYQKNSVTFGSNKL 427
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1099 KYDELQEKLSSTSEKVEAEKVHVHEKTIHLEHREHSGLE-----L 1141
QY 428 QVKENKNLEIKQHNGVIT-----BIGDNAPRNVDFQNKLRKYDLEEVKLPSTIRK 481
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1142 QSAEERSRAEINELREVLETVAQEAETDLKEKLVSLTENKVLGNEALKEGELDTK 1201
QY 482 IGAFAPQNNLKSFEASDDLEIKGAFMNNRIETLEKDKLVTTIGDAAFHNIHIAVL 541
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1202 VAMF-----DELQK-----QFSSTHAEKEBAEAKLAVHERTISHLTEVHTRSL 1244
QY 542 PESVQIEGSAFRQANNLIWFGSKVTLGEMAFLSNLEHLDLSEKQLTEIPVQAFS 601
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1245 -----ELHSAEBSKNEE-----IESKLHAELEMA-AQKEAEVKLDSKLDALIEBLGYE 1293
QY 602 DNALKEVLLPASLKTITREAPKKNHLKOLEVASALSHIAFNALDNDGDQPDNKKV-VVK 660
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1294 EQATEAARAEETHKIKPDEAVIK--IKSLEEQAVT-----ENKVELFH 1335
QY 661 THNNGYALAD-----GEHFIVDPDKLSSTIVDL-----EKILKIEGLDYSLR-Q 705
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1336 TEKENLVANSKLNELHQNKLQNELQVALAAVAEKEGSEETHSRLKRTLDGMQRKE 1395
QY 706 TTQTFQDMTTAGKALLSKSNURQEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKAT 765
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1396 ELESQVSTVESHEELKSKYNTITLEKQ--LLNEKYESAKEGELGAIKLEEQMNVKSE 1453
QY 766 KNGQL--LERSINKAVLANNSAIKANKVKLEKLDLITGLVEGKPLAQATWVGVL 823
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1454 KELHLSKLERQITLSELKTMEE-IQTMQVETTEKEALTTMQE-----HANLVH----- 1502
QY 824 LKTPLEPPEYIGLVYFDKSGKLIYALDMSDTIGSQKDAYGNPILNVDED-NRGGYHAL 882
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1503 -----BKALEQLLEVRKELNDAYH-- 1523
QY 883 AVATLADYEGLDIKTILNSKLSQLSIRQVPTAAVHR--AGIFQAIQNAABAE 934
```

Db 1524 ---TIANQE-----EQSVREIKWDAYKKFSESDQLEAEQOHVAELE 1561

RESULT 28
ABU42253
ID ABU42253 standard; protein; 3533 AA.
XX AC ABU42253;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #27780.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Staphylococcus aureus.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohleen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WI: 2003-029926/02.
XX DR N-PSDB; ACA46123.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 70177; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained

Best local similarity	15.26; 1.00; 0.021	15.26; 1.00; 0.021	15.26; 1.00; 0.021	15.26; 1.00; 0.021
Matches	231; 231; 231	231; 231; 231	231; 231; 231	231; 231; 231
Conservative	169; 169; 169	169; 169; 169	169; 169; 169	169; 169; 169
Mismatches	424; 424; 424	424; 424; 424	424; 424; 424	424; 424; 424
Indels	380; 380; 380	380; 380; 380	380; 380; 380	380; 380; 380
Gaps	52; 52; 52	52; 52; 52	52; 52; 52	52; 52; 52

SQ Sequence 1992 AA;

Query Match . 4.1%; Score 206.5; DB 6; Length 1992;
Best Local Similarity 19.0%; Pred. No. 0.0025;
Matches 220; Conservative 157; Mismatches 420; Indels 361;

[illegible]

[illegible]

Qy	165	LP	SOA	ADQ	TQ	L	I	O	V	A	S	F	A	P	T	P	D	K	T	A	E	A	Y	T	S	R	A	G	E	I	S	O	L	-----	D	V	D	G	K	E	I	215																		
Db	7729	Q	T	S	N	I	N	E	D	T	S	Q	O	V	N	F	D	E	Y	T	D	R	G	N	I	V	A	E	O	T	N	P	N	S	P	T	N	I	A	D	K	I	T	E	A	K	N	D	L	H	G	Y	Q	K	L	7788				
Qy	216	N	E	G	E	V	F	N	S	Y	L	L	K	V	T	-	I	P	T	G	K	H	G	I	Q	A	F	V	D	N	K	N	I	A	E	V	N	-----	L	P	S	L	E	T	I	S	D	Y	A	F	266									
Db	7789	K	O	A	O	Q	S	I	N	T	I	N	O	M	T	G	L	N	O	A	K	E	Q	L	N	O	E	I	Q	O	T	R	S	E	V	H	V	I	N	K	A	O	A	L	D	S	M	T	L	R	O	S	I	T	7848					
Qy	267	A	H	L	A	K	I	Q	I	D	L	P	N	I	K	A	I	G	E	L	A	F	F	O	N	-----	Q	I	T	G	K	L	S	P	R	O	L	M	R	L	A	E	R	A	F	K	S	N	H	I	K	T	E	320						
Db	7849	D	E	H	E	V	K	Q	T	S	N	Y	I	N	-	E	T	V	G	N	O	T	A	Y	N	N	A	V	D	R	V	K	I	I	N	O	T	S	N	P	T	M	P	L	E	V	E	R	A	-----	T	S	N	V	K	7902				
Qy	321	F	R	G	N	L	K	V	I	G	E	A	S	F	O	D	N	D	L	S	O	L	M	P	-----	D	G	L	E	K	I	E	S	A	F	T	G	-----	N	P	G	D	D	H	Y	N	R	V	369											
Db	7903	I	S	K	D	A	L	H	-	G	E	R	E	L	N	D	N	K	S	T	F	A	V	N	H	L	D	N	L	N	O	A	K	E	A	L	T	H	E	I	E	O	A	T	I	E	O	A	T	I	V	7960								
Qy	370	V	L	M	T	K	S	G	-----	N	P	S	G	L	A	T	E	N	T	Y	N	V	N	P	D	-----	K	S	L	W	O	E	-----	S	P	E	I	D	T	Y	K	W	-----	S	P	E	I	D	T	Y	K	W	407							
Db	7961	A	L	N	D	M	K	K	L	D	I	V	A	Q	D	N	V	R	S	N	Y	I	N	E	D	S	T	P	O	M	Y	N	D	T	I	N	H	A	O	S	I	I	D	O	V	A	N	T	G	H	D	E	-	8019						
Qy	408	L	E	E	D	F	T	Y	Q	K	N	S	V	T	G	F	S	N	K	-	G	L	O	K	V	K	R	N	K	N	L	E	I	P	K	H	G	N	G	V	T	I	E	I	G	D	N	A	F	R	N	D	F	O	N	K	T	L	R	466
Db	8020	I	E	N	A	I	N	N	I	K	H	A	I	N	A	L	D	G	E	H	K	L	O	A	K	E	N	A	L	-----	L	I	N	S	L	N	D	L	N	A	P	O	R	A	I	N	L	V	N	8070										
Qy	467	K	Y	D	I	E	E	V	K	L	P	S	T	I	R	K	I	G	A	F	A	P	Q	S	N	N	L	K	S	F	E	A	S	D	D	L	E	I	K	E	G	A	P	M	N	R	I	E	L	K	D	K	L	V	T	I	526			
Db	8071	E	A	O	T	R	-----	K	V	-----	A	E	O	L	S	A	O	A	L	N	D	A	M	K	H	L	R	S	I	Q	N	Q	-----	S	S	V	R	Q	E	S	K	Y	I	N	A	8114														
Qy	527	G	D	A	-	A	F	H	I	N	H	I	Y	A	I	-----	V	L	P	S	V																																							

RESULT 37

ADO84851

ID ADO84851 standard; protein; 10203 AA.

XX

AC ADO84851;

XX

DT 29-JUL-2004 (first entry)

S epidermidis surface anchored LPXTG protein SeqID23.

LPXTG; cell wall-anchored surface protein; Gram positive bacterium; extracellular matrix molecule; sequence database; C-terminal; immunoglobulin-like fold region; Ig-like fold region; antibacterial; vaccine; gene therapy; infection; medical device; prosthesis; premature newborn; AIDS; debilitated cancer; bone marrow transplantation.

Staphylococcus epidermidis.

WO2004025416-A2.

25-MAR-2004.

15-SEP-2003; 2003WO-US028789.

13-SEP-2002; 2002US-0410303P.

(TEXA) UNIV TEXAS A & M SYSTEM.
(INHI-) INHIBITEX INC.
(UABR-) UAB RES FOUND.

Hook M, Xu Y, Sillanpaa JV, Sthanam N, Ponnuraj K, Patti JM; Hutchins JT, Hall A; WPI; 2004-315684/29.

Identifying LPXTG-containing cell wall-anchored surface proteins from Gram positive bacteria, for treating infection caused by the bacteria, comprises searching sequence information database for the sequence having LPXTG-motif.

Claim 16; SEQ ID NO 23; 96pp; English.

This invention relates to a novel method of identifying LPXTG-containing cell wall-anchored surface proteins from Gram positive bacteria that bind to an extracellular matrix molecule which comprises searching a database of sequence information for a putative protein sequence having the LPXTG-motif in its C-terminal region and analysing the sequence for the presence of one or more Immunoglobulin (Ig)-like fold regions. The invention may be useful for the production of compounds with an antibacterial activity or for production of a vaccine. In addition the disclosed sequences may be useful for gene therapy. The antibody is useful for treating or preventing an infection of Gram-positive bacteria in a human or animal patient. The method and the proteins are useful in generating antibodies for treating and preventing the spread of infections of Gram positive bacteria, for interfering with, or inhibiting binding interactions by Gram positive bacteria, for monitoring the level of Gram positive bacterial antigens, or antibodies recognising the antigens in a human or animal patients suspected of containing the antigens or antibodies, in preventing or reducing infection of medical devices and prosthesis caused by such organisms, and in treating or preventing infections in highly susceptible groups such as premature newborns, AIDS and debilitated cancer patients, and bone marrow transplantation. The present sequence is that of a surface anchored LPXTG protein identified using the method of the invention.

Sequence 10203 AA:

RESULT 37

ADO84851

ID ADO84851 standard; protein; 10203 AA.

XX

AC ADO84851;

XX

DT 29-JUL-2004 (first entry)

```
Db 7729 QTSNVINEDTSQVNFDFYTRGKIVAVQTNPNMSPNINIAIKITEAKNDLHGQKL 7798
Qy 216 NEGEVFNYSLLKVT-IPYGVKHIQDQAFVDMKNIAEVN-----LPSELETISDYAF 266
Db 7789 KQAQQSINTINOMTGLNAQKEQLNQBIIQOTQTRSEVHQVINKAQAALNDSNMTLQSGIT 7848
Qy 267 AHLALKQDLDLNDLKAIGELAPFN-----QITGKLSLPQLMLAERAFKSHIKTIE 320
Db 7849 DEHEVQKTSNYIN-ETVGNQTAAYNNAVDVRKQIINQTSNPTMNPLEVERA-----TSNVK 7902
Qy 321 PRGNSLKVIGEASFODNDLSOLMLP---DGLKIESEAFPG-----NPGDDHYNRV 369
Db 7903 ISKDALH--GERELNDKNSKTFVAVNHLNLAQKEALTHEIQATVTSQVNNIYNKAK 7960
Qy 370 VLMTKSGK-----NPSGLATENTYVNP-----KSLMQE-----SPEIDYTKW 407
Db 7961 ALNNDMKKLKDIQAQDQNVRSNNYINEDSTPQNNYNDTINHAQSIIDQVANPTMSHDE- 8019
Qy 408 LEEDFTYQKNSVTGFSNK-GLQKVRGNKLEIPKOHNGVTTIEGDNAFRVDFQNKTLR 466
Db 8020 IENAINNIKHAINALDGEHLKQAKENANL-----LINSLDLNAPOQDAINRLVN 8070
Qy 467 KYDLREVKLPTIRKIGAFAPQSNLKSFEASDDLLEETKEGAFMNNRIETLELKDCLVTI 526
Db 8071 EAQTRÉ-----KV-----AEQLQAQALNDAMKHLRNSIQM--SSVRQBSKYINA 8114
Qy 527 GDA-AFHINHIYAI-----VLPEVQEIGRSAFRQ-----NGANNLIFMGSKVKTIGEMA 575
Db 8115 SDAKKEQYNH--AVREVENIINEQHPITLDEIKIQLTGVQVQANNDL-----8159
Qy 576 FLNSRLEHLDESEQQLTEIP-----VOAFSDNALKE-----VLLPAS 613
Db 8160 ---NGVELLDADKQNAHOSIPTLMLNQAQNALNEKINNNAVTRTEVAAIIGQAKLLDHA 8216
Qy 614 LKTIREEAFKNNHLKQLEVASALSIAFNAL--DDNDGDBQFDPN-----KVUVKTHNSY 666
Db 8217 MENLESIKDKQEVKQ-----SSNYINEDSDVQETDYNADVHTBIINQVNVPTL 8266
Qy 667 ALADGEHFTVDPKLSSITVDLEKILKLEGLD--YSTLRQTQTQFDRMTTAGKALLSK 724
Db 8267 SIEDIHAINEVNQAKQLRGQKLYQITDLADKELSKLDDLTSQSSSISNQIYVAKTR 8326
Qy 725 SNLRGQ-EKQKPLQBAQFPLGRVLDLKAIAKBAKALVTKATYNGQLLERSINKAVLAYN 783
Db 8327 TEVAQAIEKAKSLNHAM-----KALNKVYKNADKVLDSRRFINEDQPEKAYQQAINHVD 8381
Qy 784 NSAIKKAN-----VKLEKELDLTLGLVEKGKPLAQ-----ATMVGVVLLKTPPLP 831
Db 8382 SIIRHQTNPEDPTVINSTHELETAQNNLHGDQQLAHAAQQDAANVING--LIHLNVQR 8439
Qy 832 EYVGLNVYFDKSGKLIYALDMSDTIGEG-----QKDAYGNPILN-----VDEDNEGYHAL 882
Db 8440 EVMINTNTNATREKAVKNDNAQALDRAMELTQQVVAHKNNILNLSKYLNEDESK-YQQQ 8498
Qy 883 AVATLADVEGLDKITLNSKLSQLSIRQVPTAAVHRAGIFQAIQNAABAEQLLPKPGT 942
Db 8499 YDRVIADEAQL-----LNQTTN-----PTLEPYKVDIVK--DNVLANEKILF-----8538
Qy 943 HSEKSSSESANSKDRGLQSNPKTRGRHS 972
Db 8539 GAELKSYDKS--NANDEIKHMYLNNAQKOS 8567
RESULT 38
ID ADS04803
XX ADS04803 standard; protein; 10203 AA.
AC ADS04803;
XX ADS04803;
DT 04-NOV-2004 (first entry)
XX Staphylococcus epidermis polypeptide seqid 4098.
DE
```

```
XX
KW antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW recombinant expression vector; infection; computer readable medium;
XX computer based system.
OS Staphylococcus epidermidis.
XX US2004147734-A1.
XX 29-JUL-2004.
XX 01-DEC-2003; 2003US-00724972.
XX 08-NOV-1997; 97US-0064964P.
XX 13-AUG-1998; 98US-00134001.
XX 29-NOV-1999; 99US-00450969.
XX (DOUC/) DOUCETTE-STAMM L.
XX (BUSH/) BUSH D.
XX Doucette-Stamm L, Bush D;
XX WPI; 2004-580138/56.
XX N-PSDB; ADS01031.
XX New isolated polypeptide and encoding nucleic acid derived from
XX Staphylococcus epidermidis, useful for diagnosing, preventing and/or
XX treating an S. epidermidis bacterial infection.
XX Claim 17; SEQ ID NO 4098; 741pp; English.
XX The invention describes an isolated nucleic acid comprising a nucleotide
XX sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
XX 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
XX of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
XX given in the specification. Also described are: a recombinant expression
XX vector; a cell comprising a recombinant expression vector of (1);
XX producing an S. epidermidis polypeptide; an isolated nucleic acid
XX comprising a nucleotide sequence of at least 8 nucleotides in length; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection, comprising a nucleic acid cited above and a carrier; treating
XX a subject for S. epidermidis infection; a recombinant or substantially
XX pure preparation of an S. epidermidis polypeptide or its fragment; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection; detecting the presence of a Staphylococcus nucleic acid in a
XX sample; a computer readable medium having recorded in it the nucleotide
XX sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
XX system for identifying fragments of the Staphylococcus genome of
XX commercial importance; a computer based system for identifying fragments
XX of the Staphylococcus plasmids of commercial importance; identifying
XX commercially important nucleic acid fragments of the Staphylococcus
XX genome and/or plasmids; and identifying an expression modulating fragment
XX of the Staphylococcus genome and/or plasmids. The methods and
XX compositions of the present invention are useful for the diagnosis,
XX prevention and/or treatment of an Staphylococcus epidermidis bacterial
XX infection. This is the amino acid sequence of a S. epidermis protein of
XX the invention.
XX Sequence 10203 AA;
XX
Query Match 4.0%; Score 205; DB 8; Length 10203;
Best Local Similarity 19.1%; Pred. No. 0.031;
Matches 201; Conservative 178; Mismatches 447; Indels 224; Gaps 45;
Qy 47 YAESGSKSLKINETSGPV--DDTVTLDFSKRTTPKIKONLAKGPREQELKAVTENTE 104
Db 7618 YNQAVNAKKNININDOPTFVMADEIQSVLNEVKTF----KQNL-----HGDQKLANDKTD 7668
Qy 105 SEKQITSGSQLEQSKESLSLNKTVPTSTNWEICDPTKGNLTVLGSLSGVKEKLSQTHLV 164
Db 7669 AQATLNALNYLNAQAGNLETKVQNSNSRPEVQVQLANQLNDAMKLLDALTGNDAIK 7728
Qy 165 LPSQAADGTQLIQVASFAPFTDKKTAETYSRAGEISQL-----DVGDKI 215
```



```
QY 295 GKLSPLQLMRLAERAFKSNHIKTIIEPRGNLSKVIGESPDNDLSQMLPDLGLEKIESE 354
Db 1531 EIL-----NNNETNEKGNDIGPV-RAAYE-----EGLNINAA 1564
QY 355 APTGNPGDDHNNRVLVWTSGKNPSGLATENTYVNDPKSLMQESPEIDYTKWLEEDFTY 414
Db 1565 TTTGD-----VTTAKDTAVQKVQQLHANPVKKPAGK-----KELDQAAAD 1604
QY 415 QKNSVTGFSNGKLGQVKRNKNLEIPKOHNGVTITEIGDGNFRVDFQNKTLKYDLEEVK 474
Db 1605 KKTQIEGTNPASQOEINDAKQ-EVDTELN-----QAKTNVD-QSSTNEYVD----- 1648
QY 475 LPSTIRKIGAFQSNLKSFEASDDLEIEKEGAFMNNRIETLE--LKOKLVITIGDAAHP 532
Db 1649 -----NAVKEGKAKINAVKTFSEYKQOALAKIEDA--- 1678
QY 533 INHIYAIVLPESVQETGRSAFRONGANNLIFMGSGVKVTIGEMAFNLNRLEHL-----DL 586
Db 1679 -----YNAKVNEADN-----SNASTSSEIAEAKQKLAELKQTADQNV 1715
QY 587 SEQQLTEIPVOAFSD-NALKEVLLPASLK-----TIREAFKK 624
Db 1716 NQATSKDDIEVQIHNLDNINDITIPGKESATTDLYAYADQKKNISADTNATQDEKQ 1775
QY 625 NHLKOLE--VASALSHIAFNALDNDGDE-----QPDNKKVVKTHHNSYALA- 669
Db 1776 QAIKQVQNVQTALENIN-NGVDNGVDVDDALTOGKAADITQVDAITVKKANQAIKAEKAE 1834
QY 670 DGBHFIVDPDKLSS--TIVDLEKILKIEGLDYSTLRQTTQTFQDMTITAGKALLSKSNL 727
Db 1835 DTKESIDHSDQLTAEBKTEALAMIKQITDQAKQGITDATTAEVEKAKAQGLEAPDNIQI 1894
QY 728 ROGEKOKFQEQAFQIFGR---VDLDKATAKAEKALVT-----KKATNGOLLERSIN 776
Db 1895 DSTEKQKALIELETALDQIEAGVNVDAADATTEEKEAFTNALEDILSKATED--ISDQTTN 1952
QY 777 KAVLAYNNSAIIKANVKRL-----EKELDLLTGLVEGGLPAAQATMVQ 820
Db 1953 AETATVNSALEQLKQKRNPPVKKVALEAIRVNVKQIEII-----KNADADASAKS- 2005
QY 821 VYLLKTPLEPEYIIGLVYFDK-SGKL-----IYAL-----DMS 854
Db 2006 ---IART-----DLGRYFDRFADKLDKQNTQNTVEAELQNVITIPAEIAIVPQNDPDAN 2054
QY 855 DT-IGEGQKDAYGNPILNVDEDEGHALAVATLADYEGLDIKTLNLSKLSQTSIRQVP 913
Db 2055 DTNNGTNDNDATANSNANATPENTGQPNVSETT--DNGKADASPTTPNNSDAATGETTVT 2112
QY 914 TAAVHRAGIFQAIQNAAEAEQLLP-----KPGTHSEKSSSES 952
Db 2113 SATDDAKDFQANNSSADASTNSPTMDNDVTSKPEVESTNNGTTDKPVTETDNATPAES 2172
QY 953 A---NSKORGLOSNPKNRGRHSAILPRTGS 980
Db 2173 TTNNNSTTTATNENAPTG---STATAPTTAS 2200
```

RESULT 41

ABU15887
ID ABU15887 standard; protein; 6713 AA.

XX AC ABU15887;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #1414.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Staphylococcus aureus.

XX PN WO200277183-A2.

XX XX

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (SLIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA19757.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 43811; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC of the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 6713 AA;

XX Query Match 4.0%; Score 203; DB 6; Length 6713;

XX Best Local Similarity 19.5%; Pred No. 0.023;

XX Matches 232; Conservative 183; Mismatches 396; Indels 376; Gaps 64;

QY 34 KQTQASSISGAD-----YAESGSKSLKINETSGP-VDDTDTDLFSDKRTTPEKIDN 86

Db 5050 QQTQAGSKFINEDKPKDAYQAAVQNAQXDLINQTNPTLDKSVQQLTQAVTT---AKDN 5106

QY '87 L---AKGPREQELKAVTEN-----TESEKQITSSQLEQSKS 120

Db 5107 LHGDQKLRDQQAQAVTTVTNALPNLHAAQQTLDINAAPTTRTEVAQHVQTATELDAME 5166

QY 121 SLSLNKTVTPSTNWEICDFITKGNLTVLGSKSGVEKLSQTDHLVLPQAADGTQLOIVAS 180

Db 5167 TLK-NKV-----DQVNTDKAQNPTTEASTDKKAVD-----QALQAAQ 5203
QY 181 FAFTPD-----KKTAAEYTSRAGENGESQLDQVDGKEIINEGEVFNYSLLKKTPTPTGY 235
Db 5204 SITDPTNGSNANKDAVEQALTKLQB--KVNEL--NGNERNVABEAKTQAKQTIDQLT----- 5254
QY 236 KHIGQDAFVDKN-----IAEV-----NLPESLETISDYAFALHALKQ-IDL----- 276
Db 5255 -HLNADQIATAKONTDOATKLOPIAELVDQATQLNQMSDLQAAVNEHANVEQTIDYTOA 5313
QY 277 -PDNLKAIGE-LAFFDN--QITGKLSLPLROLML--AERAP-----KSNHIKITE 320
Db 5314 DSDQKQAYKQALADAEVNLKQANKQOQVDOALONTLNKAKQALNGDERVALAKTNGKHID 5373
QY 321 FRGSLKVIGEASF-----QNDLSQLM-----LPDGLKIESEAFPTGNPGDDHYNN 367
Db 5374 -QLNALNNAQQDGFGRIDQSDNLQIQIIVDEAKALNRAMDQL--SQEITGNEGR----- 5426
QY 368 RVVLWTKSGKNPSGLATENTYVNPDKSLWQESPE-IDYTKW-----LE 409
Db 5427 -----TKGSTN-----YVNADTVQKVYDEAVDKAKQALDKSSGQNLTAEQVIKLN 5472
QY 410 EBDFTVQKNSVTGFS-----NKGI-QKVKR 432
Db 5473 DAVTAAKKALNGEERLNKRBALQRLDQLTHLNNAQRLAQIQINNAAETLNVASRAINR 5532
QY 433 NKNLE-----IPKONGV-----TITEIGDNAPRV-----DFQNKTLRK 467
Db 5533 ATKLDNAMGAVQYIDEGHLGVISSTNYINADDNLKANYDNAIAAHELDKVGQNAIAK 5592
QY 468 YLEEVK-----LPSTIRKIGAFQSNL-----KSFASDDLEB 503
Db 5593 ABAEQLKQNIIDQNALNGDQNLAKKANAFAVNSLNGLNQOQDQLAHKAINNADTVSD 5652
QY 504 IKEGAPMNRIBTLELKLKLVIGDAAFHINHIYAIPLPESVQIEGRSAFRONGANNLIF 563
Db 5653 VTD--IVNNQID-----LNDAMETLKLHVDNEIPNAEQTVNYQNADNNAKTNFDD 5700
QY 564 MGSVKYTL-----GEMAFLSNRLEHLDLSEQKQ-LTEIPVQAFSD---NALKE 607
Db 5701 AKRLANTLLNSDNTVNDINGAIAQVNDAIHNLGDQRLQDAKDKAIOISINQALANKLKE 5760
QY 608 VLLPASLKTIREAPKQNLHQLQEVASALSHTAFNALDNDGDQFD---NKVVVKTTHN 664
Db 5761 I--EASNATDQDKLAKKAEBEL-ANSIINNIT--NKATSNQAVSQVQTAGNHAIBQVHAN 5815
QY 665 SYVALDGEHFIVDPDKLSSTIVDLBK-ILKLTIEGLDYS-----TLRQTTQ 710
Db 5816 EIPKAK-----IDANK-----DVKQVQALLIDEIDRPNLTDKQKALKDRINQILOQG 5864
QY 711 PRDMTTAGKALSKSNLRGGEKQ--KFLQEAQFFL-GRVLDLDAKAKAEKALVTK----- 762
Db 5865 HNDINNA---LTKEIEQAKAQAQALQDIKDLVKAKEDAKQDVQKQVQALIDEIDQNP 5920
QY 763 -KATNGQLLERSINKAVLAYNNSAIKANVVRLEKELDLLTGLVEGKGPLAQATWQGV 821
Db 5921 NLTDKEQALKDRINQ--ILQOHHGGINNAWTK---EIEE-----QAKAQAQA----- 5964
QY 822 VLLKTPPLPEYIYIGLVYVFDKSGK-LIVALDMSDTIGEGOKDAYGNPILNVDEO--NEG 878
Db 5965 --LKEIKDLVKAKENAKQDVQKQVQALIDEIDQNPNLTDKEQKALKDRINQILOGHNDI 6022
QY 879 YHAL-----AVATLADYEGH-DIKTLNLSKLSQTSIRQVPTAAYHRAGIFQAIQNA 930
Db 6023 NNAMTKEBIEQAKAQA--QALQDIKDLVKAKEDAKNAIKAL-----AN 6064
QY 931 ABAEQLLKPKGTHSEKSSSS--ESANSKDRGLQS--NPKT-----NRG 969
Db 6065 AKRQDQINGNPDLTPBQKALKAEIDEAEKRALQNVENAQTIDQLNRG 6111

RESULT 42
AAU34207

ID AAU34207 standard; protein; 2025 AA.
XX AAU34207;
AC
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #483.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Staphylococcus aureus.
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX N-PSDB; AAS52066.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Example 3; SEQ ID NO 5703; Slipp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes,
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2025 AA;

Query Match 4.0%; Score 202.5; DB 4; Length 2025;
Best Local Similarity 18.5%; Pred. No. 0.0046;
Matches 216; Conservative 174; Mismatches 410; Indels 367; Gaps 55;

QY 16 SVVTHNQEVFLVKEPILKQTOASSISGADYAESS--GKSKL-----KINET 61
Db 413 AAVOHAKDLINQTSNPTLDKQAVQEQLTQGVNQAKNLHGDQKQADDDKQHAVTDNLQNSL 472
QY 62 SGPVDDTVTDLFSDKRTTPPEKIKDNLAKGPP-EQELKAVTENTSEKQITSGSQLEQSK 120
Db 473 NNPPQALLESQINNAATFDE-VAQKLAELAQALDQAMQALNRNSIQDQQQTSSGKPF----- 526

Db 1604 NNPOQALQSINNAATRDE - VAKLABAQAALDQAWQALRNSIQDQQQTSGSKF----- 1657
Qy 121 SLSLNKTVSTNWEIC-----DFTYK-ONTVLVGLSKSGVEKLSQT-----DHL----- 163
Db 1658 ---INEDKPKDAYQAQVQAHKDLINOTGNPT--LDKQVEQLTOAVTTAKDNLHGDKQL 1712
Qy 164 -----VLP-----SQAADGTQLIQVASFAPTPDKKTAIAEYTSRAGE----- 200
Db 1713 ARDQQAQVTVNALPNLNAHQOQALUDALNA-----PRTREVAQHVTATDELHAMET 1766
Qy 201 -NGEISOLD-----VDGKEIINEGEVFNYSYLLKKVTIPTGYKHIGQDAFVD-- 245
Db 1767 LKNKYDQVNTDKAQPNYTEASTDKKEAVDQA---LQAAESITDPTGNSNANKAVEQAL 1822
Qy 246 -----NKMAEYNLPESLETISDYAPHL-----ALKQIDLPDLNKAIGEL 286
Db 1823 TKLOBKVNELNGERNVABAK-AAQAKQTIDQ--LAHLNADQIATAKONIDQATKLOPIABL 1879
Qy 287 AFFDNQI-----TGKLSLPROLMRLAERAPKSHIK-- 317
Db 1880 VDAQTLQNSMDQLOQAQVNEHANVETQVDYTDQSDSKQNAKQAIABEAENVLKQNSNQ 1939
Qy 318 -----TIEFRGSLKVIQBSFQDNDLSQMLPDGLEKTESBAFTGNPGDDHYN 366
Db 1940 VDAQLQNLNAQAALNGDERVALAKTNGK-HDIDQL---NALNNAQDQGFGRIDQSHDL 1995
Qy 367 NRVVLWTKSGK-----NPSGLATENTYNNPKSLWQESPE-IDYTKW--- 407
Db 1996 NQIQOIVDEAKALNRAMDQLSOEISGNEGRTKSGTNYVNAQTQVQVYDEAVKAKQALD 2055
Qy 408 -----LEEDFTYQKNSVTG--PSNKGLOKVKR-----NKNLETPK 440
Db 2056 KSTQNLRTABQVILKNDVATAAKKALNGEERLNKRSBALQRLDQLTHLNNAQRLALQ 2115
Qy 441 QHNGVTITEIGD--NAPRNVDPQNKTLRKY-DLEEVKLPSTIRKIGAFAFOSNNLKS--- 494
Db 2116 INNAETLNKASRAINRATKLDNAMGAVQYIDEQHLGVISSTNYNA---DDNLKANYD 2171
Qy 495 ---FASDDLEIKGAPFNNRIETLELKDKLVITIGDAAFHINHIYAVLPESVQEIORS 551
Db 2172 NAIANAHELDKVOGNATA--KABEALQKQNIIDAQNAL----- 2208
Qy 552 AFRQGANNLIPMGSKVKTGEMAFSLNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLP 611
Db 2209 -----NGDQNLNAKOKAN-----AFVN-----SLNGLMQ----- 2233
Qy 612 ASLKTIREAPFKNHLKOLEVASALSHIAFNALDNDGDEQP-----DNKVVVYKTHNSYA 667
Db 2234 -----QQDLAHNAINNADTVSDVDIVNNQIDLDNAMEDTLKHLVDNEIPNAEQTVNYQ 2287
Qy 668 LADGEHPIVDPKLSSTIVDLKILKLEGLDYSTLRQTQTQFRDMTTAGKALLSKSNL 727
Db 2288 NAD-----DNAKTNFDDAKRLA-----NTLNSDNTNVNDINGAIQAVKDAIQN 2331
Qy 728 ROGKQKFLQEAQFFLGRVLDLKAIAKAEKALVTK-----KATNGQLLERSINKA-V 779
Db 2332 LNGEOR--LOEAK-----DKALQNVNKLADKLEIASNATDQKLIK--NKABE 2379
Qy 780 LAYNNSAIKANVRLEKELDLITGLVEGKGLPQAQATVQGVYLLKTLPLPEYVIGLVN 839
Db 2380 LA--NSIIN--NINKATSNQDVSQVQTAGNQAEQVHANE-----IPKAKIDANK 2425
Qy 840 YPKSGK-LIYALDMSDTIGEQKDAYGNPILNVDEB--NEGYPHALAVATLADYEG----- 892
Db 2426 DVDKQVQALIDIEIGNPNLTDKEKQALKDRINQILQQGHNDINNALTKEAIEQAKERLAQ 2485
Qy 893 --LDIKTLNLSKLSQLTSTROVPTAAHYRAGIFQAIQNAAAAEQLPKPGTHSKSSSS 950
Db 2486 ALQDILKDLVAKEDAKNKIKAL-----ANAKROQINSNPDLTPQKAKA 2529
Qy 951 --ESANSKDRGLQS--NPKT-----NRG 969
Db 2530 LKEIDBAEKALQNVENAGTIDQLNRG 2556

RESULT 44

AAV19935

ID AAV19935 standard; protein; 1087 AA.

XX AC AAV19935;

XX DT 19-JUL-1999 (first entry)

XX DE B. burgdorferi antigenic protein, t742.aa.

XX KW Antigenic protein; vaccine; Lyme disease; infection; detection.

XX OS Borrelia burgdorferi.

XX PN WO9859071-A1.

XX PD 30-DEC-1998.

XX PF 18-JUN-1998; 98WO-US012718.

XX PR 20-JUN-1997; 97US-0050359P.

XX PR 22-JUL-1997; 97US-0053344P.

XX PR 22-JUL-1997; 97US-0053377P.

XX PR 03-SEP-1997; 97US-0057483P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (MEDI-) MEDIMMUNE INC.

XX PI Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX DR WPI; 1999-189980/16.

XX DR N-PSDB; AAX61632.

XX CC New isolated Borrelia burgdorferi nucleic acids - used to develop

XX CC products for the diagnosis, prevention and treatment of diseases caused

XX CC by Borrelia, particularly Lyme disease.

XX PS Claim 12; Page 125; 275pp; English.

XX CC This sequence represents a Borrelia burgdorferi (Bb) protein of the

XX CC invention, which is suitable for use in a vaccine. The Bb polypeptides

XX CC can be used in vaccines for eliciting protective antibodies to members of

XX CC the Borrelia genus, particularly for the use against Lyme disease in

XX CC humans and animals. They can be used for preventing or attenuating an

XX CC infection caused by a member of the Borrelia genus. The products can also

XX CC be used for detection of members of the Borrelia genus

SQ Sequence 1087 AA;

Query Match 3.9%; Score 200; DB 2; Length 1087;

Best Local Similarity 19.6%; Pred. No. 0.0027;

Matches 212; Conservative 164; Mismatches 358; Indels 347; Gaps 54;

Qy 18 VTHNQEVPSLVKEPIILKOTQASSISGADYAESGSKLKLINETSQVDDTVTDLFSDKR 77

Db 156 VNDQKNLFNLEK---LKK-----NLGKSNSENILNDSQKIE-----NDKQ 193

Qy 78 TT---PEKIKDNLAGPREQE-----LKAVTENTESEKQITSGSQLEQSK-----ES 121

Db 194 NTNLSKEKNSENILKTPDNSKYNNNNNTTSLKKTSSNSQKESLSPSPQTIIGKIYRPS 253

Qy 122 LSLNKTVPSTSNWEICDFITKNTLVG-----LSKSGV-EKLSQTHLVLPQAADGTQ 174

Db 254 YLIKKEL-----YBILDDINTGRVTGLGNRLKELIKGLSNKFKQVKNELIENSKNEASN 308

Qy 175 LI-----QVASFAPTPDKKTAETYSRAGENGEISQDVGKEIINEGEVFNYSY 224

Db 309 LLLTLIKKIDIEPNLINIPKPYK-----EIFQDKEDKK-----POY 346

Qy 225 L-----LKKVTIPTGYKHIGQDAFVD-----NKMAEYNLPESLETI 261

Db 347 LEDLSKSVHSIKPIDLENTKSR--QQAIKDLNEFLKNNPNDQAQSKTLAQANKIQHLEDL 404
Qy 262 SDYAFALHAKQIDLPDNLK----AIGELAFF-----DNQITGKLSLPRQLMRLAERAF 311
Db 405 KSKVH---SIKPIDL-ENTKSRQQAIDKLNFLKNNPNDQAQSKTLAQANKIQHLEDLS 460
Qy 312 KSNHKTIEF-----RGNLSKVIGRASPOND-----LSQLMPLDGLKIESEAPT 357
Db 461 KVHSIKPIDLENTKSRQQAIDKLN--FLKNNPNDQAQSKTLAQANKIQHLEDLSKVHS 518
Qy 358 GNPQDDHYNNRVVLWTGSKNPSGLATENTYVNPDKSLWOESPEIDYTKWLEEDTYQKN 417
Db 519 IKPID-----LENT-----KSRQQAIDKLN-----EFX-KNN 544
Qy 418 SVTGFSNKGLOKVRKNKLEIPKQHGVTITEIGNAFRNVDFQNKTLRKYDLEEVKLPS 477
Db 545 PNDQAQSKTLAQANKIQHLEDLSK-----VHSIKPIDLENTKSRQQAIDKLN--- 592
Qy 478 TIRKIGAFQSNLKSFEASDDBEIKEGAFMNNRIETLE-LKDKLVITIGDAAPHINHI 536
Db 593 -----EFXKNNPNDQAQSKTLAQ-----NKIQHLEDLSK-----VHSI 627
Qy 537 YAVIL-----PESQVEIGRSFRQNGANNLIFMGSKVKTIGEMAFLSNRLHDLSEOK 590
Db 628 KPIDLENTKSRQQAIDKLN--NEFLKNNPND-----AQASKTLAQ-----ANKIQHLEDLSK 677
Qy 591 QUTEIPVOAFSDNALKEVLLPASLKTIREAFKNNHLKQLEVASALSHTAFNALDNDGD 650
Db 678 VHSIKPIDLENTKSRQQAIDKLN--AIKDLNE--FKNNPNDQAQSKTLAQ--AY-----ENNGD 725
Qy 651 -----EQFDNKVVVKT-----HNN-----SYALADGEHFI-----VDPDKLSS----- 683
Db 726 LLKAENAYEKIILNTQSDHYKLGIIIRPKLKYEHSIESFDQTIKLOPKKKKALHNKI 785
Qy 684 TIVLEKILKLEGLDYSTLROTTQTFRDMTTAGKALLSKNLROGEKQKFLQBAQ--- 740
Db 786 ALMLNKNKKAIESFE-----KAIQIDKNYGTAVYQKGIABEKNGD 826
Qy 741 -----FFLGRVDLQKAIKAEKALVTKKATKNGQLLESIN-----KAVLAYNNS 785
Db 827 MQQAFASFKNAYLNDRKNPNYALKAGIVSNLGNFKQSEYLNFFNANAKPNEIAIYNLS 886
Qy 786 AIKKANVRLEKELDLTLGLVE-----GKPLAQATWQGVYLLKTP 828
Db 887 IAKFEN-NKLESLETINKAIDLNEPKSEYLYLKASINLKKNYQNAISLYSVIEKPN- 944
Qy 829 PLPEYIYGNVYFDKGLIYALD-MSDTIGSGQDAYGN----- 867
Db 945 ENTSAYINLAKAYERSGKNSQAISTLEKIINKNNKLNALNGLYKKEKNYQKAIETFEK 1004
Qy 868 PILNVDENEGYHALAVATLADYEGLDIKTLNLSKLSQTSIROVPTAAYHRAGIFQAIQ 927
Db 1005 AIIN--SDIEAKYNLATTLEINDNTRAKDLR-----EYTKLPNNPEALHALGIEYNE 1058
Qy 928 N 928
Db 1059 N 1059

RESULT 45
ID AAY19934
XX AAY19934 standard; protein; 1119 AA.

AC AAY19934;

XX 19-JUL-1999 (first entry)

DE B. burgdorferi antigenic protein, f742.aa.

XX Antigenic protein; vaccine; Lyme disease; infection; detection.

OS Borrelia burgdorferi.

PN WO9859071-A1.
XX PD 30-DEC-1998.
XX PF 18-JUN-1998; 98WO-US012718.
XX PR 20-JUN-1997; 97US-0050359P.
XX PR 22-JUL-1997; 97US-0053344P.
XX PR 22-JUL-1997; 97US-0053377P.
XX PR 03-SEP-1997; 97US-0057483P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
WPI; 1999-189980/16.
DR N-P8DB; AAX61631.
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases caused
PT by Borrelia, particularly Lyme disease.
XX Claim 12; Page 124-125; 275pp; English.
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus
XX SQ Sequence 1119 AA;

Query Match 3.9%; Score 200; DB 2; Length 1119;
Best Local Similarity 19.6%; Pred. No. 0.0028;
Matches 212; Conservative 164; Mismatches 358; Indels 347; Gaps 54;

Qy 18 VTHQEVSLVKEPIKQTOASSISGADYAESKSKLKNETSGPVDVDTDFSDKR 77
Db 188 VNDQKNLNFLEK---LKK-----NLSGKNSENILNDSQKIE-----NDKQ 225
Qy 78 TT---PEKIKONLAKPREQE-----LKAVENTESEKQITSGOLEQSK---ES 121
Db 226 NINLSKEKNSENILKTPDNSKYNNNTSLKTSNSQKSELSFPQTIIKLYRYS 285
Qy 122 LSLANKTPSTSNWEICDPIITKNTLVG-----LSKSGV-EKLSQTDHLVPSQAADGTQ 174
Db 286 YLIKKEL-----YEILDDINTGRVTLGKNNRLKELIKGLSNKFQKVNELIENSKNEASN 340
Qy 175 LI-----QVASFAFTPDKTAIAEYTSRAGENCEISQLDVGKEINEGEVFNYS 224
Db 341 LLLTLIKKIDBPNLINIPKPYKK-----EIFQDKEDKK-----PQY 378
Qy 225 L-----LKKVTIPTGYKHIGODAFVD-----NKNIAEYNLPSELETI 261
Db 379 LEDLSKSVHSIKPIDLENTKSR--QQAIDKLNFLKNNPNDQAQSKTLAQANKIQHLEDL 436
Qy 262 SDYAFALHAKQIDLPDNLK----AIGELAFF-----DNQITGKLSLPRQLMRLAERAF 311
Db 437 KSKVH---SIKPIDL-ENTKSRQQAIDKLNFLKNNPNDQAQSKTLAQANKIQHLEDLS 492
Qy 312 KSNHKTIEF-----RGNLSKVIGRASPOND-----LSQLMPLDGLKIESEAPT 357
Db 493 KVHSIKPIDLENTKSRQQAIDKLN--FLKNNPNDQAQSKTLAQANKIQHLEDLSKVHS 550
Qy 358 GNPQDDHYNNRVVLWTGSKNPSGLATENTYVNPDKSLWOESPEIDYTKWLEEDTYQKN 417
Db 551 IKPID-----LENT-----KSRQQAIDKLN-----EFX-KNN 576
Qy 418 SVTGFSNKGLOKVRKNKLEIPKQHGVTITEIGNAFRNVDFQNKTLRKYDLEEVKLPS 477

Qy	776	NKAVLAYNSAIKANVAKRLEKDLITGL--VEGKGPLAQATMVQGVYLLKTPPLPPEY	833
Db	1227	NKELL---NKDIRNKDLKISIGNMEQOQTGLKNTPSKGQ-----QNTGLKNTP----	1270
Qy	834	YIGLVVYFDSGGKLIYALDMSDTTIGEQOKAYGNPILNVDEDNEGYHALAVATLADYGL	893
Db	1271	-----NENQOQTG-----LKNTPSEGOOQTGLKNTPSEGOOQTG-----	1304
Qy	894	DIKTILMSKLSQLTSIIRQVPTAAVYHRAGIFQAIQNAAAEAEQLLPKPTHSEKSSSESA	953
Db	1305	-LKNTPNR--QOQTGLKNTPSEGOOQTG----LKNTPIEGQOQTGLKNTPSEGOOQTGLK	1358
Qy	954	NSKDRGLQS---NPKTNRGRHSAILPRTGSKG	982
Db	1359	NAANKGOQTGLKNAANKGOOQTGLKNTPSKG	1390

RESULT 47	
AAW98149	
ID	AAW98149 standard; protein; 2285 AA.
XX	
AC	AAW98149;
XX	
DT	05-JUL-1999 (first entry)
XX	
DE	Bacillus subtilis metalloprotease YOMI.
XX	
XX	YOMI; metalloprotease; protease; textile; animal feed; detergent;
KW	Gram-positive bacterium.
KW	
XX	
OS	Bacillus subtilis.
XX	
PN	WO9914342-A1.
XX	
PD	25-MAR-1999.
XX	
PF	08-SEP-1998; 98WO-US018828.
XX	
PR	15-SEP-1997; 97GB-00019636.
XX	
XX	(GEMV) GENENCOR INT INC.
PA	(GEMV) GENENCOR INT BV.
XX	
PI	Estell DA;
XX	
DR	WPI; 1999-229541/19.
DR	N-PSDB; AAX24980.
XX	
XX	Compositions containing metalloprotease from gram-positive microorganism.
PT	Claim 4; Fig 1A-O; 59pp; English.
XX	
PS	

The present sequence is metalloprotease YOMI of *Bacillus subtilis*. The invention relates to the discovery of this previously unknown metalloprotease, uses of the metalloprotease in industrial and agricultural applications, and advantageous strain improvements based on genetically engineering a Gram-positive microorganism to delete, underexpress or overexpress the enzyme. Due to overall relatedness of the enzyme with *Pseudomonas* lasA protein, YOMI appears to be a member of the M23 metalloprotease family. The metalloprotease can be used in claimed cleaning compositions, animal feed and compositions for the treatment of textiles. It may also be used for peptide hydrolysis, waste treatment and for cleaving recombinant fusion proteins. Expression vectors comprising a nucleic acid (see AAX24980) encoding the metalloprotease and host cells are claimed. Gram-positive cells in which YOMI is inactivated (by gene mutation or deletion) are used for production of heterologous proteins, especially enzymes, hormones, growth factors and cytokines.

Sequence 2285 AA;

Query Match 3.9%; Score 199.5; DB 2; Length 2285;
Best Local Similarity 20.9%; Pred. No. 0.0082;
Matches 215; Conservative 155; Mismatches 356; Indels 303;

Qy	81	EKIDNIAKPREBELKAVNTES--EKQITSSQLEBSQESISLN-----	125
Db	771	KKAKDDFEQS--QQTNNVAITTKNDSTDKLIQQYKELQVKRESLTSDEBQEVLTQVQOL	829
Qy	126	-KTVPSITSNWEICDPIFKGNTLVGLSKSGVEKLSQT--DHLVLPQAAA-----	DGQT 174
Db	830	AQTPPAL-----VKYDSQGNAILKTNKELSKALENTKEYILALKKQETRDSAKTTFEDASK	885
Qy	175	LIQVSPAFTPDKKTAIAEYTSRAGENGESQLDVGKEIINBGEVENSYLLAKVTIPTG	234
Db	886	EIKKSDELQYKQ--JADYNDKGRPKWOLIIDDDYKVAADRAK--QSMLKQAGSDIESG	941
Qy	235	YKHIGQDAFVDNKNIAEVNIPESLET--ISDYAFALHALKOIDIDPNLKAIGEIAFPNQI	293
Db	942	NAKYKDSVLSTANAYSSIDISNTLKTISD-----VVNKLNLKDDLDLP--EELEKFFSSSL	994
Qy	294	TGKLSLPQLMRLAERAFKSNHIIKTIEFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIES	353
Db	995	-GKLQERKQ-----KALDSGDEKAPD--NAKK-----DUOSLL-----	1024
Qy	354	EAFGTGPDGHHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKWLEBDF	413
Db	1025	-----ETYSKSDSS-----IDVFKMSFD--K 1043	
Qy	414	YQKNSVTGFNGKLGQVKVRNKULIEIPKHNGVITIEIGDNAFNRVDPQNTLARKYDLEEV	473
Db	1044	AQRNIKDG--DKSLSSVKSEVG-----DLGETLAEAGNEA--EDPGKK-----LKEA	1086
Qy	474	KLPTIRKIGAPAFQSNNLKSFEASDDLEIEIKGAFMNNRIETLELKKLVTIGDAAPHI	533
Db	1087	LDANSVDDIIKAAIKEMSDAMQF--DSVDVLNGDIFNN-----TKDQVAPLND-----	1132
Qy	534	NHIVAIPLPESVOBIRGSAPRQGANLIIPMGSKVKTGLGMAFLNRLHLDLSEQKOLT	593
Db	1133	-----LLEKMAE-GKS-ISANEANTLI--QXKELAQAIISIENGCVKINRDEVIKOR	1180
Qy	594	EIPVQAFSD-----NALK-----EVLPLASLKTIRBE-----	620
Db	1181	KVKLDAYNDMVTYSNKLMTKEVNNAIKTLNADTLRIDSLKKLKERKLDMSAEALSDEV	1240
Qy	621	-----AFKQHLKQLE-----VASAL--SHIAFNALDDNDGDQDF	654
Db	1241	KSINNADADAKELKKLEKMLQPCGYSNSQIEAMQSVKSALESYI--SASEEATSTQEMN	1298
Qy	655	NKVVVKTHNSYALADGEHFIVDPKLSSTIVD-----LEKILKLIEGL-----DYSTLR	704
Db	1299	QKALVEAGTSLUENWDOQEKANEBETKTSMTVDVKYKEALKEVNAEIDKYNQVNDYPKYS	1358
Qy	705	QTTOTQPRD--MTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLAKIAKAEBKALVTKK	763
Db	1359	Q-----KYRDAIKKEIKALQOKKLMQ--EQAKLLKD-QIKSGNI-----TQYGI	1404
Qy	764	ATKNG--QLLERSINKAVLAYNNSAIKKANVKELEKELDLTLGLVEGK--GPLAQATWVQG	820
Db	1405	TSSGGTSPSTGSGYSYSSYINSAAKYNV-----DPALIAAVIQESGFNAKARSGVG	1459
Qy	821	VYLLKTPLPYPIYIGLVNVPFDSKGLIYALDMSDITGEGQK--DAYGNPILNVDDEDNEG	879
Db	1460	AMGLMQMPATAKSLGVNNAYDP-----YQNVMGGTQYLQOOLEKFGF--NVEKILAAY	1511
Qy	880	HALAVATLADYEGLD-----IKTIL-----NSKLSQITSIRQVPTAAYHRAGIOAION	928
Db	1512	NA--CPGNVIKYGIPPPKETQNVYKKTIMANYKSLSSATS-----SIASY-----	1556
Qy	929	AAAEAEQLPKPGTHSKSSSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG--SFVY	986
Db	1557	-----TNNSAFRVSSKYGOESGLASSP--HKGTDFAAKAGTAIKSLQSGKV	1601
Qy	987	GILGYTSA 995	
Db	1602	QIAGYSKTA 1610	

95

1602 QIAGYSKTA 1610

XX Vancomycin; teicoplanin; MRSA; diagnosis; SA1577; FmtB;
 KW antibiotic resistance.

OS Staphylococcus aureus.

XX WO2003062466-A2.

XX 31-JUL-2003.

XX 16-JAN-2003; 2003WO-GB000129.

XX 16-JAN-2002; 2002GB-0000865.

XX (UYBR-) UNIV BRISTOL.

XX Walsh TR, Avison MB, Howe RA;

XX WPI; 2003-636743/60.

DR N-PSDB; ACF05836.

XX
 PT Detecting a strain of *Staphylococcus aureus* having reduced susceptibility
 PT to vancomycin and/or teicoplanin, comprises detecting the presence of
 PT mutation(s) leading to loss of function of at least one gene.

XX Claim 26; Fig 1; 78pp; English.

XX
 CC The present sequence is that of a putative FmtB-like protein encoded by
 CC open reading frame SA1577 of methicillin resistant *Staphylococcus aureus*
 CC (MRSA) strain N315. SA1577 has been assigned the provisional gene name
 CC Mrp on the basis of homology of its sequence to known genes. This gene is
 CC mutated in vancomycin resistant *Staphylococcus aureus* strain Mu50, the
 CC mutation leading to truncation of the encoded protein (see ABB62800). On
 CC the basis of such loss of function mutations, the invention provides
 CC polypeptides, polynucleotides, probes, primers, antibodies, kits and
 CC methods for detecting strains of *S. aureus* having reduced susceptibility
 CC to vancomycin and/or teicoplanin. The methods are rapid, reproducible and
 CC sensitive, and allow for detection of strains that are fully or
 CC heterogeneously resistant

XX Sequence 2196 AA;

Query Match 3.9%; Score 199; DB 7; Length 2196;

Best Local Similarity 19.2%; Pred. No. 0.0084;

Matches 217; Conservative 182; Mismatches 387; Indels 342; Gaps 58;

QY 20 HNOEVSILKEPIKQTOASSISGADYAESSGSKLKINETSGPVDVTDLFSD---- 75

DB 1207 YNAKLAENATPDATNDEKNAINTLNDRQQAIESIKQANTNAEVDQAATVAENNIDAV 1266

QY 76 -----KRTTPEKIKDNLAKGREQLKAVTENTESEKQITSGSOLEQSK-----SLSLN 125

DB 1267 QVDVVKQAARDKITAEVAK--RIEAVKQTPNATDEEKQ-AAVNQINQLKDAQAINQINQN 1323

QY 126 KT---VPSTSNWEI-----CDFITKGNITVLGSKSGVEKLSQTDHLVPSQAADGTQL 175

DB 1324 QTNQDQDVTITNAVNDAIDNVEAEVVIKPKAIADIEKAVEKQOQIDNSL---DSTNEK- 1379

QY 176 IQVASFAFTPKKTAETAETYSRAGEISQLDVGKEIINEGEVNSVLLKVT--IPT 233

DB 1380 -EVASQALAKEKEKALAA-IDQAQTSQVNQAATNGVSAIK-----IIQETKVKA 1429

QY 234 GYKHIGDQAVDNKNTAEVNLPE-----SLETISDYAFALAL-----KQID 275

DB 1430 ARKINQKA---NELRAKINQDKAEATBERQVALDKINE--FVNQAMTDTNRTNQVD 1484

QY 276 LPDNLKAIGELAFPDNQITKLSLPLRLMLAER-----AFKSNHIKTIETPRGNSLKV 328

DB 1485 -DITSQALDSIA-----LVAPHHVRAARDKAVKQYEAKEIEQAEHATDEEQ 1534

QY 329 IGEASFQDND-----LSQMLPDGLEKIESEFTGNPGDDHNNRVLTWTKSGNPSGL 382

DB 1535 VALNQLANNEKLALQINQAVTNDVKRVETNGIATLKG---VQPHIVI-----KPEAQ 1586

QY 383 ATENTYVNPDKSLWOESP-----EIDYTKWL-----BEDFTYQKNSVTFGSN--- 424
 DB 1587 AIKATAENQVESI-KDTPHATVDELDEANQLISDILKQAQQOIEINTNQDAAVTDVNRQTI 1645
 QY 425 KGLQKVK-----RNKN-----LEIPKOHNGVITTEIGD--NAFRNVD 459
 DB 1646 KAIEQIKPVRKRRAALDSIENNNKQLDAIRNTLDTTQDERDVAIDTLNKIVNTIKNDI 1705
 QY 460 FONKTLRKVDLEE-----VKL-----PSTIRKIGAFQFSNNILKFEASD-DLEEI 504
 DB 1706 AQKNTNAEVDRTETGDNINIKVILPKVQVKKPAAROSVGKAEQAQNAL--IDOSDIESTE 1763
 QY 505 KEGA-----FMNRIETLELKLVTIGDAAFINHIYAVILP-----ESVQETG 549
 DB 1764 RLAAKHLVEQALNQALDQINHADKTAQVNNQDSIDAQNIISKIKPATTVKATALQIQNTA 1823
 QY 550 RSAFRONGANN-----LIFGSKVK---TLGEMAFI-----SNRLEHLD 585
 DB 1824 TNKINLIKANNEATDEQONIAIAQVEKELIKAKQOIASAVTNADVAYLLHDEKNEIRETE 1883
 QY 586 -----LSEQKOLTEIPQAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSHI 639
 DB 1884 PVINKASAREQLTTL-----FND---KKQAIENAIQATVEE--RNSILAQLQ---NIYDT 1931
 QY 640 AFNALDDNDGDQDFDN--KVVKVTHNSYALADGEHFIVDPDKLSSITVDLEKILKIEG 697
 DB 1932 AIGQIDQDRSNAQVNDKTASLNIQTIDHL-----DVPIKPPDAEKTINDDLARVLTALVO- 1985
 QY 698 LDYSTLRQTTQTPRDMTTAGKALLSKNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEK 757
 DB 1986 -----NYRKVSNRNK-----ADALK 2000
 QY 758 ALVTYKATKNGOLLERSIN---KAVLAYNNSAIKANVKRLEKELDLLTGLVEGKPLAQ 814
 DB 2001 AITALKLQWDEELKTARTNADVDVLRKFNVALSIEAVITEKNSLLR--IDN---IAQ 2055
 QY 815 ATMWQGVYLLKTPPLPEYVIGLVYFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDE 874
 DB 2056 QT-----YAKPKAIATPQLAKVKVLIQO-----YVAD-----GNRM--IDE 2090
 QY 875 DNEGHALAVATLADYEGDLITKILNSKLSQITSIRQVPTAAYHRAGIFQAIONAAAEAE 934
 DB 2091 D-----ATLN-----DIKQHTQFIVDEILAIAK-LPAEA-----TKVSPKEI 2125
 QY 935 QLLPKPGCTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG 982
 DB 2126 QPAPKVCTPIKKBETHESRKVEKE-----LPTGSEG 2157

RESULT 52

ABB63519

ID ABB63519 standard; protein; 2346 AA.

XX

AC ABB63519;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 17349.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.
 XX PA Venter JC, Adams M, Li PMD, Myers EW;
 XX PI WPI; 2001-656860/75.
 XX DR N-PSDB; ABL07622.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX PS Disclosure; SEQ ID NO 17349; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB1616176-AB1630511), expressed DNA
 CC sequences (AB161840-AB161875) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX CC
 XX SQ Sequence 2346 AA;
 Query Match 3.9%; Score 198.5; DB 4; Length 2346;
 Best Local Similarity 18.2%; Pred. No. 0.0098;
 Matches 218; Conservative 181; Mismatch 396; Indels 403; Gaps 45;
 QY 6 KTVALT-----TTVSVTHNQVFLVKEPILKQTOASSISGADYAESGSKL 56
 DB 225 KTESLKMGEQYEAQVKTIGELTSKIE---MQNDTAFKQNAQTEYVGLKKELDAKEL 281
 QY 57 KINETSQVDDVTDLF---SDKRTTPKIKDNIAKGR-----EOLKAVTENTSEK 107
 DB 282 -----FEIFKSTSDHLIQRELLQGEISEIKRLLEBAEQCAQTEQMETWK 328
 QY 108 QITSGSQLESLSLNTKPTSTNWEICDPTKNTLVGLSGVGEKLSQTDHLVLP 167
 DB 329 QKXSAELDEONKIQAMEQELASND-----LLKQAESNLESALCOLAP 373
 QY 168 QAADGTQIQVAFAPTPPKTAIABYTSRAGE-----NGEISQLDVGKEINE----- 217
 DB 374 SAAVASRLIR-SDLSLT-----ELYSMYAKSSELEMRNCEIEQLKQLKLSIIAIESAP 428
 QY 218 -----GEVFNLSYLLKVTIPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAF 266
 DB 429 ILEKQNSDYQKKMETNSLELRE-----HDELLQNKLCLELELERALSTLNHQN 477
 QY 267 AHLALKQI-----DLP-----DNLKAIGELAF 288
 DB 478 ENKLLQTHDLSRQVCMMLDELNCIRAGVKHRIQPTQLPTSSLSISDNLVTFSSIEE 537
 QY 289 FNOQTGKLSLPRLMRLAERAFKS-----NHKTIIEFRGNSLVKTEASFPQND 338
 DB 538 LVDRNTYLLNMGRELTELLEASERKQDKMLBQSKNHRKLDARFAELEDL--LTQKNT 595
 QY 339 LSQLMLP-DGLSKIESEAFTPGDDHYNNRVVLMTKSGKPNFSGLATENTYVDPKSLWQ 397
 DB 596 VTTLSKCDRYKKLFVAA-----QKKLQNTVDLDDSNLEPNDALDTS 639
 QY 398 ESPEIDY--TKWLEBDFTYQKNSVTGFSNKGKLVKRNKNLEIPKQHNGVITTEIGDNAP 455
 DB 640 EQPAANFESRKLKVRVRLEQLEG-----EVKKYASLK----- 674
 QY 455 RVNDFQNTKLRKYD-LSEVKLSSTTRKIGAFAPQSNLKSFEASDDLEIKEGAFWN--- 511
 DB 675 ENYDYVTSKRNDALAQEQFDSMRKEVRELT--SSNCKLMNTTEFQKEIQLLHKNTGT 732
 QY 512 --NRIETLE-----LKDKLVITGDAAFHINIYAIVLVPESVOEIGR 550

Db 733 YKQVTTLEERTKNYEKTIKHEQTVHLLKDEM-----AAHRKHAADAQAQSIQRQNR 787
 QY 551 -----SAFRQNGANNLIIPGSKVKTIGEMAFISNRLHLDLSEQKOLTRIP 596
 DB 788 ILRDTSSRLQIETKETHREQOSQL-----LNSLEFIKTNLSEMEGRQRL----- 835
 QY 597 VQAFSDNALKEVLLPASLKTIRREAFK-----KNHLKOL-----EV 632
 DB 836 -EQLDDTVRE--LAAQRHRHFOESEEKPREINSFKQAETAIKLDEEKKLADKMAQEL 892
 QY 633 ASALSHIAFNALDNDGDQPDNKVVKVTHNSVALAD--GEHFIVDPDKLSSTIVLEK 690
 DB 893 TSVREELAENVKYNELSKLQEVLTPTLNDNPITAAKGRAREFEL---KLDQATVEIES 949
 QY 691 ILKLI-----EGLDYSTLRQTQTFQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGR 745
 DB 950 LTKSLAKTRHEGEQFYKMSQSAESEIKRLHGHGELVAK-----QBEETIKLRSSE----- 1000
 QY 746 VDLDKATAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKKANVKRLEKELDLLTGL 805
 DB 1001 AELKTRISDLAEAWLSNVTQSKTVNQS-----GQLKSAQDDLKSL 1042
 QY 806 VEGKGLAQATMVGQVLLKTPPLPEYIIGLVYFDKSGKLIYALDMSDTIGEGQKDAY 865
 DB 1043 LE-----KLTRANCTIRLSENTSLVESLNAEV-----KY 1074
 QY 866 GNPILNVDENEGTHALAVATLADYEGLDIKTIINLSKLSQTSIRQVPTAAHYHRAGIPQA 925
 DB 1075 ANGMIQ-----HSADIQLTRYKAEFFKA--NDELINQLKSGRESLQAAY-----DE 1118
 QY 926 IQNAAAEAEQLPKPGTHSEK-----SSSSE 951
 DB 1119 LLRSNAEAQKLDKERESEKRVADLHALNSLHDQIEALASKLAVLASQSONPNSSLNE 1178
 QY 952 SANSKDRGLQSNPKT--NRGRHSAILPRTGSGSVFYGILGYTSVALLSLITAIIKKKK 1007
 DB 1179 SAMDQDSLNASGLTAAEBGRNE-----QLLKIIFLKEK 1215
 RESULT 53
 ABR53116
 ID ABR53116 standard; protein; 1790 AA.
 XX AC ABR53116;
 XX DT 20-JUN-2003 (first entry)
 XX Protein sequence #SEQ ID 1097.
 XX Multiprotein complex; eukaryote; drug target; diagnosis.
 XX Saccharomyces cerevisiae.
 XX EP1258494-A1.
 XX PD 20-NOV-2002.
 XX PF 20-DEC-2001; 2001EP-00130253.
 XX PR 15-MAY-2001; 2001EP-00111774.
 XX PA (CELL-) CELLZOME AG.
 XX PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
 PI Marzioch M, Schultz JD, Superti-Furga GD;
 XX WPI; 2003-250078/25.
 DR N-PSDB; ACC61158.
 PT New isolated protein complexes useful for diagnosing a disease or
 PT disorder, or as a target for an active agent of a pharmaceutical,
 PT preferably a drug target in the treatment or prevention of disease or
 PT disorder.

XX PS Disclosure; SEQ ID NO 1097; 17pp + Sequence Listing; English.

CC The invention relates to multiprotein complexes from eukaryotes. Proteins of the invention and DNA sequences encoding them are given in records ABR2568-ABR53903 and ACC6010-ACC61944 respectively. The complexes are obtainable by using a protein as a bait and isolating the set of proteins which is attached thereto from cells. Such protein complexes may comprise up to 30 distinct proteins. Protein complexes of the invention are useful for diagnosing a disease or disorder, or as a target for an active agent of a pharmaceutical, preferably a drug target in the treatment or prevention of a disease or disorder. Note: the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office. The complete document is available on CD-ROM

XX SQ Sequence 1790 AA;

Query Match 3.9%; Score 197.5; DB 6; Length 1790;
Best Local Similarity 21.0%; Pred. No. 0.0078;
Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

QY 25 FSLVKEPILKOTQA--SSISGADYAESGSKLKNINETSQGVDDTVTD----- 71
DB 434 FDLQDFLLKQVLQNNSTNNVGDNAKNGSKSDKSDKDTGDKDGTGYESFKAN 493

QY 72 LF-----SDKRTTPEKI-----KONLAKGPREOELKAVTENTSEKQITSG 112
DB 494 LFEVLLNYDAELNLPFKLFFTTDIPMPFFQDQ-----KYSELREITRN-----VTYG 543

QY 113 SLEQSKESLSLNTKVP-----STS-----NWEICDFITKGNLTVLGSKS 152
DB 544 NDLE-DEEPLKAIQITISELLTTSMTAADIPIISYLTFLIYWLFGDFKATNDFL--SDKS 600

QY 153 GYEKLSQTHLVLPQAAD-----GTQLIQA-----SFAFTPDKTAIAEY--TSRAGEN- 201
DB 601 VIKSLLSFSYQI---QDEVTIKCLVTLMLGVAYEFSSKESPPRKEYPEFITKLGDKN 657

QY 202 --DETSQDQD--GKEIINEGEV-----PNSYLLKKVT-----IPTGYKH 237
DB 658 YASRIKQPKDSYFKVDNEDSILTPDELDEGLPKVYESTYFIQFNENIYRIITALSH 717

QY 238 IQQDAFVDNKNIAEVLNPE-----SLETISDYAFAHALKQIDLDPNLKAIGEL 286
DB 718 DPDERPINKISPEEVEKLQROCTKLGBEITSLQTESHTENLTKLIALTNEHKELDEK 777

QY 287 AFQDN-----QITGKLSLQRLMRLAERAPKSHNIHTIEFRGNSLKV 328
DB 778 YQILNSSHSLKENFSILETELKNVRDSLDEMTQLRDVLETKDKENQALLEYKSTIHK- 836

QY 329 IGASQFQDNDLSQMLPDGLEKIEGAFTGNPDHNNRVLVLTGSKGNPSGLATENTY 388
DB 837 -----QESIK--TLEKLETLISOK-----KKAEDGIN-----KMGKDLFALSREMOA 878

QY 389 VNPD--KSLMQESPEIDY-----TKWLEEDFTYQKNSVTGFSNKGLOKVK--RNKNLEIPKQ 441
DB 879 VEENCNKLQKXDNVNHQKETSLSKEDIAAKITEIKAI--NENLEEMKIQCNNLSKEKE 937

QY 442 HNGVITITEGD-----NAPRVNDFONTKRLKYDLEEVKLPSSTR- 480
DB 938 HISKELVEYKSRFQSHNDLVAKLTKLKLANNYKDMQAEENSLIK-AVEESKNESLIQL 996

QY 481 -----KIGAFAFQSNLKL-----SFEAS-----DDLEEIKEGAPMNNRIETLE----- 518
DB 997 SNLQNKIDSMQKFNQTERGSIKNIQLEKKTISDLEQTKKEIISKSDSKDEYSGOI 1056

QY 519 --LKDCLVITIGDA--APHINHIAVLVPESVQIGRSAPFQNGANNI--IPMGSKVKTLCGEM 574
DB 1057 SLLEKLELTATTANDENVNKISLTKTRELEAEALAAVK-NLKNLETKLSTSEKALKEV 1115

QY 575 AFLSNRLEHLDISEQKQLTEIPVQAFSDNALKEVL-----LPASLKTIREFAFKKNHLK 628
DB 1116 KENEHLEKBEKIQLEKEATETKQQLNSLRANLESKEHEDLAAQLKKYEEQIANKERQY 1175

QY 629 QLEVASALSHIAFNALDNDGDGEQFDNKVWVVKTHNSYALADGEHFIV----- 676
DB 1176 NEEISQ-----LNDEITSTQENESIKKNDDE-----LEGEVKAMKSTSEESQNLKK 1222

QY 677 -DPDKLSSTIVDLEK-----ILKLIEGLDYSTL-----RQTQTQPRDMTT 716
DB 1223 SEIDALNLQIKELKKKNETNBASLIESIKSVESETVKIKELQDECNFKEVSELEDKLK 1282

QY 717 ACKALLSKSNLRQGEKQFLQ-----EAQFFGRV-DLDKAIKAEKALVTKKAT--- 765
DB 1283 ASEDKNSKYLELQKSEKIKBEIDAKTELKIQLEKITNLKAKESSELSRLKKTSS 1342

QY 766 --KNG-QULERSINKAVLAYNNSAIKANVRLEKELDLITGLVKGKGPLAQATWVQGVY 822
DB 1343 ERKNAEQLEKLNKNE--IQIKNQA-----FEKERKLN--EGSSTITQ----- 1381

QY 823 LLKTPPLPPEYVIGLVNVPDSSGKL-----IYALDMSDTIGEGQKDAYGNPILNVDENE 877
DB 1382 -----EYSEKINTLEDELIRIQENENELKAKEIDNTRSELEKVSLSNDELSEKQN- 1431

QY 878 GVHALAVATLADYEGLDIKTILNSKLSOLTSTI-----ROVPTAAVHRAGIFOAIONAAAE 932
DB 1432 -----TIKSLQD--ELLSYKDKITRNDEKLLSTERNKRDLSLKEQ-----LRAAESKAK 1481

QY 933 AEQLLPKPGTHSEKSSSESA-----NSKD--RGLQSNPKTN 967
DB 1482 VEEGLKK-----LEBESSKEKAELEKSKEMMKLESTIESN 1517

RESULT 54
ADK63040
ID ADK63040 standard; protein; 1790 AA.
XX AC ADK63040;
XX DT 06-MAY-2004 (first entry)
XX DE Disease treating protein complex-derived protein #646.
XX KW protein complex; drug target; diagnosis.
XX OS Unidentified.
XX PN EP1338608-A2.
XX PD 27-AUG-2003.
XX PF 20-DEC-2002; 2002EP-00102902.
XX PR 20-DEC-2001; 2001EP-00130253.
XX PA (CELL-) CELLZOME AG.
XX PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;
PI Michon A, Leutwein C, Rick J;
XX WPI; 2003-638460/61.
XX DR N-PSDB; ADK63041.
XX PT New proteins and protein complexes from eukaryotes, useful as targets in
XX PT drug screening, or in diagnosing or screening for the presence of a
XX PT disease or disorder, or a predisposition for developing a disease or
XX PT disorder in a subject.
XX PS Disclosure; SEQ ID NO 1291; 13pp; English.
XX CC The invention relates to novel protein complexes comprising a first and a
XX CC second protein, or its derivative, fragment, homologue or variant. The
XX CC proteins are selected from given protein complexes, which are not defined
XX CC in the specification. The variants are encoded by nucleic acids that
XX CC hybridize to the nucleic acids encoding the proteins under low stringency

CC conditions. The protein complexes are useful as targets for an active
CC agent of a pharmaceutical. These protein complexes are particularly
CC useful as drugs targets for the treatment or preventing of a disease or
CC disorder. The complexes and methods above are useful in diagnosing or
CC screening for the presence of a disease or disorder or a predisposition
CC for developing a disease or disorder in a subject. These are also useful
CC in screening for a drug for treatment or prevention of a disease or
CC disorder. The molecule that modulates the amount, activity or protein
CC components of the complex is useful for the manufacture of a medicament
CC for the treatment or prevention of a disease or disorder. This sequence
CC corresponds to a protein of the invention. (Note: the sequence data for
CC this patent did not form part of the printed specification but was
CC obtained from the EPO in electronic format).

XX
SQ Sequence 1790 AA;

Query Match 3.9%; Score 197.5; DB 7; Length 1790;
Best Local Similarity 21.0%; Pred. No. 0.0078;
Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

QY 25 FSLVKEPILKQQA---SSISGADVAESGSKLKNINETSQVDTVTD----- 71
DB 434 FDLQDFLLKQVLQCNSTNNVGDNAKENGSGNSKSDKSDGDKDGTGTEYEGSPKAN 493
QY 72 LP-----SDKRTTPEKI-----KDLAKGPRQELKAVTENTSEKQITSG 112
DB 494 LFEVLNVDALNLPFKLFFTTDIPMPFFQDDH-----KYSEELREITRN-----VTTG 543
QY 113 SQLEQSKESLSLUNKTVP---STS-----NWEICDFITKGNLTVGLSKS 152
DB 544 NDLE-DEEPLKAIQIISLTTSLTAADIRIPISVLTFLYWLFGDFKATNDFL--SDKS 600
QY 153 GVEKLSQDTHLVPSQAAD-----GTQIIQVA---SPAFTPDKTAIAEY-TSAGEN- 201
DB 601 VIKSLSPSYQI---QDEVDITKLVMLGVAYEFSSKESPFPRKEYFEFTITKLKDN 657
QY 202 --GEISQLDVD--GKEITNGEV-----FNSYLLKAVT-----IPTGYKH 237
DB 658 YASRIKQPKKQSYFKVDNEDSILTPDELTKPVYFTFYIQLFNENYIRINTALSH 717
QY 238 IQQDAFVNQNAIYNLPE-----SLETISDYAFALHAKQLDLPDNKAIGEL 286
DB 718 DPDBEPINKISPEEVEKLQROCTKLKGETISLQTESTHENLTKLIALTNEHKELDEK 777
QY 287 AFFDN-----QITGKLSLPRQLMLAERAFKSNHKTIEFRGNSLKV 328
DB 778 YQILNSSHSLKENFSILETELKNVDRSLDEMTQLRDVLETKDKENQATLLEYKSTIHK- 836
QY 329 IGEASFPQNDLSQLMPLDGLKIESEAFGNPDGDDHNNRVVLWTKSGNPSGLATENTY 388
DB 837 -----QSDSIK--TLEKLETLISQK---KKAEDGIN-----KMGKDLFALSREMOA 878
QY 389 VNPD-KSLWQSPEDY-----TKWLEEDFTYQKNVTFGFSNKGLOKVK-RMKNLEIPKQ 441
DB 879 VEENCKNLQEKDKDKNVNHQETKSLKEDIAKITEIKAI-NENLEEMKIQCNLSKEKE 937
QY 442 HNGVTITEIGD-----NAFRNVDQNKTLRKDYLEEVKLPSTIR- 480
DB 938 HISKELVYKRFQSHNDLVAKLTSLANNYKMQAENESLIK-AVESKNESSIQL 996
QY 481 -----KIGAFAPQSNLX-----SFEAS-----DLEETKEGAFMNRITLLE----- 518
DB 997 SNLQNKIDSMQSEKENFOIERGSIENKIEQLKKTISDLEQTKBEIISKSDSKDYEQSI 1056
QY 519 --LKDQLVTIGDA-AFHNIHVIYVLPESVQIGRSAPFRQNGANNL-IFMGSKVKTLGEM 574
DB 1057 SILLEKLETATTANDENVNKISELTKTREELAEALAAVK-NLKNLETKLETSEKALKEV 1115
QY 575 AFLSNRLEHLDSEQKQITEIPVQAFSDNALKEVL-----LPASLKTIREAFKXHLK 628
DB 1116 KENEHLEKEEIKLEKEATETTKQQLNSLRANLESLEKSHEDLAAQLKYYEQIANRQY 1175
QY 629 QLEVASALSHIAFNALDDNDGDQDNKVVVKNTHNSVALADGEHFIV----- 676

DB 1176 NEEISQ-----LNDEITSTQENESIKKKNDE-----LEGFVKAMKSTSEBSNLKK 1222
QY 677 -DPKLESTIVDLBK-----ILKUIEGLDYSTL-----RQTTQTOFRDMWT 716
DB 1223 SEIDALNLIQIKELKKONETNEASLESISKSVESETVKIKELQDCNFKPEKEVSELEDKLK 1282
QY 717 AGKALLSKSNURQEKQKFLQ-----EAQFFLGRV-DLDKAIKAKAEKALVTKKAT--- 765
DB 1283 ASEDKNSKYLELOKESKIKKEELDAKTTTELKIQLEKITNLKSKESELSRLKKTSS 1342
QY 766 --KNG-OLLERSINKAVLAYNNSAIKXANVKRELKDLTLGLVEGKPLAQATWQGVY 822
DB 1343 ERKNAEQLKKNB--IQIKQA-----FKERKLLN---EGSSTITQ----- 1381
QY 823 LLKTPPLPPEYIIGLVNPFKSGKL-----IYALDMSDTIGEGQKDAYGNPILNVDENE 877
DB 1382 -----EYSEKINTLEDELIRLQENELKAKEIDNTRSELEKVSLSNDELLEEKQN- 1431
QY 878 GYHALAVATLADYEGLDIKITILNSKLSQTSI-----RQVPTAAYHRAGITQATONAAAE 932
DB 1432 -----TIKSLQD-EILSYKDKITRNDKLLSIEDNKRDLESLEKEQ-----LRAAQESKAK 1481
QY 933 AEQLLPKPGTHSEKSSSESA---NSKD--RGLQSNPKTN 967
DB 1482 VEEGLK-----LEBESSEKAELEKSEMMKKLESTIESN 1517

RESULT 55
ADN18933
ID ADN18933 standard; protein, 1790 AA.
XX AC ADN18933;
XX DT 02-DEC-2004 (first entry)
XX DE Bacterial polypeptide #1586.
XX KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX OS Bacteria.
XX PN US2003233675-A1.
XX PD 18-DEC-2003.
XX PF 20-FEB-2003; 2003US-00369493.
XX PR 21-FEB-2002; 2002US-0360039P.
XX PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WI WIPI; 2004-061375/06.
XX PT New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX PS Claim 1; SEQ ID NO 1586; 122pp; English.
XX CC The invention relates to a recombinant DNA construct comprising a

CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 1790 AA;

Query Match 3.9%; Score 197.5; DB 8; Length 1790;
 Best Local Similarity 21.0%; Pred. No. 0.0078;
 Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

QY 25 FSLVKEPILKQQA--SSISGADYAESSGSKLKINETSFPVDDTVTD-----71
 Db 434 PDLQDFLLKQVLCNNSTNNVGDNAKENGSGNSKSDSKDGTGDKGTEYEGSPKAN 493
 QY 72 LF-----SDKRTTPEKI-----KDLAKGPREQELKAVTENTESEKQITSG 112
 Db 494 LFEVLNLDYDAELNLPFKLFFTTDFMFFFOODH----KYSELREITRN-----VTTG 543
 QY 113 SLEQSKESLSLNTKVP--STS-----NWEICDFITKGNLVLGSKS 152
 Db 544 NDLE--DEPLKAIQIISILLTSLTAADIRIPISYLTFLLIYWFGDFKATNDFL--SKDS 600
 QY 153 GVEKLSQTDHLVPSQAAD-----GTQLIOVA--SFAFTPDKTAIAEY--TSRAGEN- 201
 Db 601 VIKSLLSFSYQI--QDEDTIKCLVTLMLGVAYEFSSKESPPRKEYFEFITKTLGXDN 657
 QY 202 --GETSQLDVD--GKEIINEGEV-----FNSYLLKKVT-----IPTGVKH 237
 Db 658 YASRIKQPKKDSYFSKVDNEDSILTPDELDTGLPKVYFSTYFIQLFNEINIRIRIALSH 717
 QY 238 IQQDAFVNKNIAEYNLPE-----SLETISDYAFAHALKQIDLPDLNKAIGEL 286
 Db 718 DPDEPINKISPEEVKQROCTKUKGETISLTQTESTHENLTKLIALTNHEKDELDEK 777
 QY 287 AFFDN-----QITGKLSLPRQLMLRAERAFKSNHIKTIKFRGNSLKV 328
 Db 778 YQILNSSHSLKENFSILETELKQNVKSDLEMTQLRDVLETKDKENQTALLEKYSYTIHK- 836
 QY 329 IGEASFQNDLSQLMPLDGLKIESEAFNGPDGHHNNRVVLTWKSGKNPSGLATENTY 388
 Db 837 -----QEDSIK--TLEKLETLISQK--KKAEDGIN-----KMGKDLFALSREMQA 878
 QY 389 VNPD--KSLWQESPEIDY-----TKWLEDPFTYQKNSVTGFSNKGKQVK--RNKNLEIPKQ 441
 Db 879 VEENCKNLQKDKGNVNHQKTKSLKEDIAAKITEIKAI--NENLEEMKIQCNLSKKE 937
 QY 442 HNGVITTEIGD-----NAFRNVDPQNTKLRYDLEEVKLPSTIR- 480
 Db 938 HLSKELVEYKSFQSHDNLVAKLTEKLKSLANNYKDMQENESLIK-AVEESKNESLQL 996
 QY 481 -----KIGAFAPQSNLKK-----SFPAS-----DLEETKEGAFMNNRIETLE-----518
 Db 997 SNLQNKIDMSQEKENFQIERGSIIEKNIEQLKTTISDLQTEKTEIISKSDSKDBYESQI 1056

QY 519 --LKDKLVITIGDA-AFHINHIYAIVLPSVOEIGRSAPFRONGANNL-IPMGSKVKTIGSM 574
 Db 1057 SILKEKLETATTANDENVNKISELTKTREELAEALAAVK-NLKNELTETKLTSEKALKKEV 1115
 QY 575 AFLNRLBHLDSQEQKQTEIPVQAFSDNALKEVL-----LPASLKTIREFAFKKNHLUK 628
 Db 1116 KENEBHLKEEIKLEKEATEKTQQLNSLRANLESLEKEHEDIAAQLKKYEEQIANKERQY 1175
 QY 629 QLEVASALSHAFNALDNDGDQEPDNKVVKVTHNSVALADGEHFIV-----676
 Db 1176 NEEISQ-----LNDEITSTQOENESIKKONDE----LEGEVKAMKSTSEESNLKK 1222
 QY 677 -DPQKLSSTIVDLK-----ILKLTGDLDYSTL-----RQTTOTQFRDMMT 716
 Db 1223 SEIDALNLIQIKELKKKNETNEASLLSISVSEITVKIKELQDCNFKKEVSELEDKLK 1282
 QY 717 AGKALLSKNLRQGEKQKFLQ-----EAQFFLGRV-DLDKAIKAEKALVTKKAT---765
 Db 1283 ASEDKNKYLELQKESEKIKEELDAKTTTELKIQLEKITNLKAKESESELSRLKKTSS 1342
 QY 766 --KNG-QLLERSINKAVLAYNNSAIKANVKLEKELDILTLVEGKGLAQATWVGQY 822
 Db 1343 ERKNAEQLKLNKNS--IQIKNQ-----FEKERKLLN---EGSSTITO-----1381
 QY 823 LLKTPPLPPEYVIGLVNLYFDKSGKL-----IYALDMSDTIGEGQKDAYGNPILNVDEDN 877
 Db 1382 -----EYSEKINTLEDELIRLQENELKAKEIDNTRSELEKVSLSNDELSEKQN- 1431
 QY 878 GYHALAVATLADYEGLDIKTILNSKLSQTSI-----ROVPTAAYHRAGIFQAIQNAAB 932
 Db 1432 -----TIKSLQD-EILSYKDKITRNDKLSLSTERNKRDLESLEKQ-----LRAAQESKAK 1481
 QY 933 AEQLLPKPGTHSEKSSSESSESA---NSKD--RGLQSNPKTN 967
 Db 1482 VEEGLKK---LEBESSKEKAELEKSKEMMKKLESTIESN 1517

RESULT 56
 AAU34139
 ID AAU34139 standard; protein; 2368 AA.
 XX
 AC AAU34139;
 DT 14-FEB-2002 (first entry)
 XX
 DE Staphylococcus aureus cellular proliferation protein #415.
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.
 OS Staphylococcus aureus.
 XX
 XX WO200170955-A2.
 XX 27-SEP-2001.
 XX
 XX 21-MAR-2001; 2001WO-US009180.
 XX
 XX 21-MAR-2000; 2000US-0191078P.
 XX 23-MAY-2000; 2000US-0206849P.
 XX 26-MAY-2000; 2000US-0207727P.
 XX 23-OCT-2000; 2000US-0242578P.
 XX 27-NOV-2000; 2000US-0253625P.
 XX 22-DEC-2000; 2000US-0257931P.
 XX 16-FEB-2001; 2001US-0269308P.
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 XX WPI; 2001-611495/70.

DR N-PSDB; AAS51998.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids.

XX

PS

PS Example 3; SEQ ID NO 5635; 511pp; English.

XX

CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the genes,

CC their use in the discovery of novel antibiotics, the essential genes

CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*

CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,

CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also

CC useful for the identification of potential new targets for antibiotic

CC development. The antisense nucleic acids can also be used to identify

CC proteins used in proliferation, to express these proteins, and to obtain

CC antibodies capable of binding to the expressed proteins. The proteins can

CC be used to screen compounds in rational drug discovery programmes. The

CC antisense nucleic acid sequence is also useful to screen for homologous

CC nucleic acids which are required for cell proliferation in a wide variety

CC of organisms. The present sequence represents an essential prokaryotic

CC cellular proliferation protein. Note: The sequence data for this patent

CC did not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 2368 AA;

Query Match 3.9%; Score 197.5; DB 4; Length 2368;

Best Local Similarity 19.8%; Pred. No. 0.011;

Matches 242; Conservative 165; Mismatches 459; Indels 359; Gaps 59;

QY 19 THNQVFLSVEPIKQTOAASSISGADYASSG-----KSKUKINETSGPVDVDTVD 71

DB 850 TQNTVSNATNEAEADAA-----VEAAQKQGLHDIQVVKSKQEVADTKSVLKDINA 903

QY 72 LPSDKRTTPEKIKO-NLAKGPREQELKAVTENTSEKQITSGSQLESKESLKNKTVPS 130

DB 904 IQTQAKVPAADTEVENATYRKQEIQNSNASTTEEKQ-AAYTELDTKKQEARNLDAAN 962

QY 131 TSNMEICDPIT-KGNTLVGLS-----KSGVEKLS----- 158

DB 963 TNS-----DVTTAKONGIAINQVQAATTKSDAKAEIAQKASERKTAIEAMNDSTTEEQ 1018

QY 159 ----QTDHLVLPQQA-----ADGTQLIQVASF-AFTPD-----KKTAI-----A 192

DB 1019 AAKDKVDQAVVTANADIDNAAANTDVDNAKTNEATIAAITPDANVVKPTAQAIADKVQA 1078

QY 193 EYTSRAGENGISQLDVGKEIINEGEVFNLSVLLKKVTIPTGYKHIGQDAFVD---NKN 249

DB 1079 QETAI DANNGATTEEKAQKQVQ-----TEKTTADTAIDGAHTNAEVEAAKNAB 1129

QY 250 AEVLNPESLETISDYAFALHALKQIDLPDLKAIGEL-----AFPDNQITGKLSLP 300

DB 1130 AKIRAIQPATTTKONAKQAIATKANERKTAQTQDITABEIAAANAVDNAV----- 1183

QY 301 RQLMLRAEPKSNHIKTIIEFRGSLKVIIGRASFDNDLSQMLPLPDGLEKTESAFTGN- 359

DB 1184 -QANNIEAANSQNDV-----DOAKTTGEASIDQ-----VTPVNNKKAATVTDKNN 1229

QY 360 ---PGDDHNNRVNVLWTKSGKN-----PSGLATENTVNPDKSLWQESPEIDVTWLEE 410

DB 1230 ITAATDD---NGVDTAKDAGKNSIQSTQPATVKSNAKNDVDQAVTTQQAIDNTGTAT- 1285

QY 411 DFTYQKNSVTGFSNKGLOKVRKNKLEIPKOHNGVTITEIGDAPRNV-----DFQNKTL 465

DB 1286 ---TEEKNAKDL-----VLKAKEKAQDILNNAQTNDVTQIKDQAVADVQGITADTTIKDV 1339

QY 466 RKYDL-----EVLKLPSTI-----RKIGAFAPQSN-NLKSFEASDDLEIEKEG 507

DB 1340 AKDELATKAREQKALIAQTADATTEBEQANQVQDAELTQGNQNIENASQSIDDVNTAKDN 1399

QY 508 APM-----NNRIETL-ELKDKLVITIGDAAFHINHIAIYVLPSVQIEG--RS 551

DB 1400 AIQAIDPIQASTDVKTNARAELELTEMQKQTEILNNNTNE-----EKGNDIGPVRA 1452

QY 552 AFRQGANNLIFMGSKVKTKLGEAMFLSNRLEHLDSKQKQLTEIPVQAFSDNALKEVLLP 611

DB 1453 AY-BEGLNNI-----NAATTTG-----DVTTAKDTAVQKVQQLHANPVKK---P 1492

QY 612 ASLTKTIREAP--KKNHLKQL-----EVASALSHIAFNALDDNDGDEQFPNK 656

DB 1493 AG-KTALDQAAADKKTOIEOTPNASQOEINDAKQEVDTLNQAKTN-IDQSSTDEYVDNA 1550

QY 657 V-----VVKT--HNSVALADGEHF-----IVDPD-----KLSSTVDL-EKTLKLEIG 697

DB 1551 VKEGAKINAVKTSFYKKKALAKIEAAYNAKVTEADNSNASTSEIAEAKQKLAELKQT 1610

QY 698 LDYSLRQLTQTQTPRDMTTAGKALLSKNSLRQGEKQKFLQEAQFFL-----GRVDLDKAIA 753

DB 1611 ADQNVNQAATSKDDIEVQIHNDLDNIINDYTIPTGKESATTDLYAYADQKKNISADTNAT 1670

QY 754 KAEKALVTKATNGQLLERSINKAV-----LAYNSAI-----KCAN----- 791

DB 1671 QDERQQAIKQVDQNVQTALESINNGVDVDDALTOGKAIDAIOVDATVYKPKANQVID 1730

QY 792 ----VKRLEKELELLTG-----LVGKGLQAQTMVQGVYLLKTLPLPLEY 833

DB 1731 AKABETKESIDQSLTAEBKTEALAMIKQITDQAKQGITDATTAEVEKAKAQ----- 1784

QY 834 YIGL-----NVYFDKSGK-----LIYALDM-----SDTIGBQKDAYGNPILNVDED 875

DB 1785 --GLEAFDNIQIDSTEKQKAIEELELTALDQIEAGVNVVDADATTE-EKEAFTNAL----- 1835

QY 876 NEGTHALAVATLADY-BGLDIKTLNLSKLSQTSIRQVPTAAVYHRAGI-----PQAI 926

DB 1836 -EDLSKATEDISDQTTNAEIAIVKNSALEBQAKQIRNPVVKNALEAIREVVNKQIEII 1894

QY 927 QNAAABAE-----QLLPKPGTHSEKS-----SSSESAN 954

DB 1895 KNADADASAKIEARTDLGRYDFRPAKLDKDTQTNTVEAELQNVTIPIAEIIVPQNDPNAN 1954

QY 955 SKDRGLQSNPKNTRGRHSAILPRTG 979

DB 1955 DTNCGSDNDNATANSNANATPENTG 1979

XX

XX AAU36796 standard; protein; 2368 AA.

XX AAU36796;

AC AAU36796;

XX

XX 14-FEB-2002 (first entry)

DT

XX Staphylococcus aureus cellular proliferation protein #966.

DE

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;

KW antibacterial; drug design.

KW

OS

XX Staphylococcus aureus.

XX

PN WO200170955-A2.

XX

PD 27-SEP-2001.

XX

XX 21-MAR-2001; 2001WO-US009180.

PF

XX 21-MAR-2000; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.

PR 26-MAY-2000; 2000US-0207727P.

PR 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.

PR 22-DEC-2000; 2000US-0257931P.

PR 16-FEB-2001; 2001US-0269308P.

XX

(ELIT-) ELITRA PHARM INC.

PA Haselbeck R.; Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX

XX WPI; 2001-611495/70.
 DR N-PSDB; AAS54655.
 XX

PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX

PS Example 3; SEQ ID NO 12389; 511pp; English.
 XX

CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

XX Sequence 2368 AA;

Query Match 3.9%; Score 197.5; DB 4; Length 2368;

Best Local Similarity 19.8%; Pred. No. 0.011;

Matches 242; Conservative 165; Mismatches 459; Indels 359; Gaps 59;

QY 19 THNQEVSLVLPKILKQTOASSISGADYAESG-----KSKLKINETSQVDDTVTD 71
 DB TQNTVSNATNEVEAEADAA-----VEAQQLGHLDIQVKSQGEVADTKSKVLDDKINA 903
 QY 72 LFSDRRTTPEKID-NLAKGPREQELKAVTENTSEKQITSGSQLESKESLSNKTVP 130
 DB IQTQAKVPAADTEVENAYNRKQEIQNSNASTEKQ-AAYTELDTKKQEARTNLDAN 962
 QY 131 TSNWEICDPIT-KGNVLVGLS-----KSGVEKLS----- 158
 DB TNS-----DVTAKNGIAAINQVQAATTKKSDAKAEIAQKASERKTAIEAMNDSTTEQ 1018
 QY 159 -----QTDHLVLPQA-----ADGTQLIQVASF-AFTPD-----KKTAI-----A 192
 DB 1019 AAKDKVDQAVVANADIDNAAANTVDNNAKTNEATIAAITPDANVKTAKQAIADKQVA 1078
 QY 193 EYTSRAGENGEISQLDVDGKEIINEGEVFNYSLLKKVITPTGYKHIGQDAFVD---NKN 249
 DB 1079 QETAI DANNGATEEKAQKQVQ-----TEKTTADTAIDGAHTNAVEAAKNAEI 1129
 QY 250 AEVNIPESETTSDFAFALALQIDLDPNLKAIGEL-----AFPDNOITGKLSLP 300
 DB 1130 AKIEAIQPAATTTKDNKAQAIATKANERKTAIAQTQDITAEETAAANANVNDNAV 1183
 QY 301 RQLMLAEAFKSNHIKTIIEPRGNSLKVIGESFQDNDLSQLMPLGLEKIESEAF 359
 DB 1184 -QANNIEAANSQNDV-----DQAKTGEASIDQ-----VTPTVKKKATAVTDKKN 1229
 QY 360 ---PGDDHYNNRWLWTKSGKN-----PSGLATENTYVNPDKSLWQSPESPIDYTKWLEE 410
 DB 1230 ITAATDD---NGVDYAKAGKNSIQSTQATAVKSNKNDVDQAVTTQQAIDNTTGAT- 1285
 QY 411 DFTYQKNSVTGFSNKGLOKVKRNKLEIPKQHNGVYITEIGDNAPRV-----DFQNKTL 465
 DB 1286 ---TEEKNAKOL-----VLKAKEKAVQDILNAQTNDVTQIKQAVADVQGITADTTIKDV 1339

QY 466 RYVDL-----BEVKLPSTI-----RKIGAFQSN-NLKSPASDDLEIKRG 507
 DB 1340 AKDELATAREQKALIAQTADATTBEKQANQVDAELTQGNQNIENAGSIDDVNTAKON 1399
 QY 508 APM-----NNRIETL-ELKDKLVITGDAAFHINHIIYAIVLPESVOEIG--RS 551
 DB 1400 ATOAIDPIQASTDVTNARAELLTEMQNKITILNNTTNE-----EKGNDIGPVA 1452
 QY 552 AFRONGANNLIFMGSKVKTLCGEMAFLSNRLHLDLSEQQLTEIPVQAFSDNALKEVL 611
 DB 1453 AY-EEGLNI-----NAATTG-----DVTAKDTAVQKVQQLHANPVKK---P 1492
 QY 612 ASLKTIREEAF--KKNHLKQL-----EVASALSHIAFNALDDNDGDEOFDNK 656
 DB 1493 AG-KTALQAAADKKTOIEQTPNASQOEINDAKQEVDTFELNQAKTN-IDQSSTDEVDNA 1550
 QY 657 V-----VVKT--HNSYALADGEHF-----IVDPD-----KLSSTIVDL-EKILK 697
 DB 1551 VKGEKAKINAVKTFSEYKDALAKTEAAYNNAKVTEADNSNASTSSEIAEAKQKLAEL 1610
 QY 698 LDYSTILROTTQTPRDMTTAGKALLSKNLRQGEKQKLEQAEQFEL-----GRVDLDKATA 753
 DB 1611 ADQNVNQAATSKDDIEVQIHNDLNDNTYIPTGKESATTDLYAYADQKKNISADTNAT 1670
 QY 754 KAEKALVTKCATKNGQLLERSINKAV-----LAYNNSAI-----KXAN----- 791
 DB 1671 QDEKQAIKQVDQNVQTALESINNGVDGVDVDDALDTQGAALDAIQVDATVVKANQVID 1730
 QY 792 -----VKRLEKELDLTG-----LVEKGPLQAQTMVQGVYLLKTPLPPEY 833
 DB 1731 AKAEETKESIQSDQLTABEKEALAMIKITDQAKQGITDATTAAVEKAKAQ----- 1784
 QY 834 YIGL-----NVYDKSGK-----LIYALDM-----SDTIGEGQKDAYGHPILNVDED 875
 DB 1785 ---GLEAFNIIQIDSTEKQKAIELETAIDQIEAGVNVVDADATTE-EKEAFTNAL----- 1835
 QY 876 NEGHALAVATLADY-EGLDIKTILNSKLSOLTSIRQVPTAAHYHRAGI-----FOAI 926
 DB 1836 -EDILSKATEDISDQTTNAETATVKNLSALEQLKAQRINPVVKNNALEAIREVVNKQIEII 1894
 QY 927 QNAABAE-----QLLPKPTHSEKS-----SSSESAN 954
 DB 1895 KNADADASAKETARTDLGRYDFRADKLDKTKTNTNEVAELQNVITIPATEAIVPQNDPNAN 1954
 QY 955 SKDRGLQSNKPNRGRHSAILPRTG 979
 DB 1955 DTNSGSDNDATANSNANATPENTG 1979
 RESULT 58
 AAB18161
 ID AAB18161 standard; protein; 2441 AA.
 XX
 AC AAB18161;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:18.
 XX
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200025728-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 05-NOV-1999; 98WO-US026796.
 XX
 PR 05-NOV-1998; 98US-0107131P.
 XX

FT Region 560. .581
/label= tandem_repeat
/note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 582. .603
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 604. .625
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 626. .647
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 648. .669
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 670. .691
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 692. .713
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 714. .735
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 736. .757
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 758. .779
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 780. .801
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 802. .823
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 824. .845
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 846. .867
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 868. .889
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 890. .911
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 912. .933
FT /label= tandem_repeat
FT /note= "one of 21 complete segments of homology of 22
FT amino acid length"
FT 934. .946
FT /label= partial_tandem_repeat
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 949. .967
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"

FT Region 968. .986
/label= tandem_repeat
/note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 987. .1005
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 1006. .1024
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 1025. .1043
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 1044. .1062
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 1063. .1081
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 1082. .1100
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 1101. .1119
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 1120. .1138
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 1139. .1157
FT /label= tandem_repeat
FT /note= "one of 11 complete segments of homology of 19
FT amino acid length"
FT 1158. .1173
FT /label= partial_tandem_repeat
FT /note= "partial segment of homology"
FT 1179. .1193
FT /label= tandem_repeat
FT /note= "one of 4 complete segments of homology of 15
FT amino acid length"
FT 1194. .1208
FT /label= tandem_repeat
FT /note= "one of 4 complete segments of homology of 15
FT amino acid length"
FT 1209. .1223
FT /label= tandem_repeat
FT /note= "one of 4 complete segments of homology of 15
FT amino acid length"
FT 1224. .1238
FT /label= tandem_repeat
FT /note= "one of 4 complete segments of homology of 15
FT amino acid length"
FT 1248. .1260
FT /label= tandem_repeat
FT /note= "one of 27 complete segments of homology of 13
FT amino acid length"
FT 1261. .1273
FT /label= tandem_repeat
FT /note= "one of 27 complete segments of homology of 13
FT amino acid length"
FT 1274. .1286
FT /label= tandem_repeat
FT /note= "one of 27 complete segments of homology of 13
FT amino acid length"
FT 1287. .1299
FT /label= tandem_repeat
FT /note= "one of 27 complete segments of homology of 13
FT amino acid length"

Query Match		3.9%; Score 196.5; DB 7; Length 1875;
Best Local Similarity		20.8%; Pred. No. 0.0095;
Matches 242; Conservative 164; Mismatches 378; Indels 379; Gaps 59;		
Qy	34	KQTQASSI--SGADYAESGSKLKINET---SGPVDVDTVDL-----PSDKRT 78
Db	199	RKTQELTLQNNNDWLEKLSK---NQYLSYRQKTDKVILDIRNELRLRNDPQWERT 255
Qy	79	TPEKIK---DNLAKGPRQ--ELKAVTENTSEKQITSG-----SOLEQSKR 120
Db	256	NNVLLKQKNNELSKLQELKEIKGLSDLSNSEQFSAEMSLKQRLVDLLESQNAVKE 315
Qy	121	SLS-----LN--KTVPSTSNWEICDPITKGNLTGLSGVSKLSQTDHLVLPQA----- 169
Db	316	ELNSIRELNTAKVITADDSKKQ---TPNEEDLLKELQTKKLAQCEKCEKRLSLSSITDEA 371
Qy	170	-ADGTQLTQVASFAPTPDKTAIABYTSRAGENGEISQLDDVGKE---IIN----- 216
Db	372	DEDNENLSAKSSDPFLKQKLIKERRTKEHLQNIETFIVELEHKVPIINSFKERTDML 431
Qy	217	EGEVNSYLLKKVTIPTGYKHIGQDAFVDNKNIAEVNLPESLETISDYAFALHALKQIDL 276
Db	432	ENELNNAALL-----LEHTSNE---KNAKVKELNAKQ-----KLVEC 466
Qy	277	PDNLKAIGELAFFDNQITGKLSLQPLMLAERAFKSNHIK-----TIEP----- 321
Db	467	ENDLQTLTK-----QRDLCKQIQYLIITNSVNSDNGPRLKBEIQIQTIONIMQEDD 517
Qy	322	---RGNLSKVIGASFQNDLSQLM-----LPDGLE-----KIESE 354
Db	518	STITESQKVVERTELVFKNIIQLQKNAELLKVVRNLADKLESKEKSKQSLOKIESE 577
Qy	355	AFTNGPGDDHNNRVLMWTKSGQNPGLATENTVYNPKSLWQSPESPIDYTKWLEEDPT 414
Db	578	-TVNEAKE-----AIITLKSEK-----MDLESRIEELQKELEELKTS 613
Qy	415	QKNSVTGFSNKGLOKVXENK-NLETPKQHNGVITIGDNAFRNVDFOKTLRKVDLEEV 473
Db	614	VPNEDASVNTIKQLTETKDLSESQVQLQTRISQITRESTENSLINKIEQ---DLYDS 671
Qy	474	KLPSTIRKIG-----AFQSNNLKSFASDDLEIKEGAFMNNRI----- 514
Db	672	KSDISI-KLGEKSSRIILAEERFKLLSNTLDLTRAEND-QLRKRPDYQLNTILKQDSKTH 729
Qy	515	-----ETLEKOKLVITIGDAAFI--NHIIYAVLPESV--QETGRSAFRONGANNLIFMGSK 567
Db	730	ETLNEYVSCSKLSIVETELLNLKBEQKLRVHLEKNLQELNKLSPK---DSLRIWVTQ 786
Qy	568	VKTIGEMAFNLRLHLDLSQKQLTEIPVQAFSDNALKEVLLPASLAKTIREEAPKKNHL 627
Db	787	LQTLQKER--BDLLEETRSQCKKIDEL-----EDALSE-----LK--KETSQKHII 830
Qy	628	KQLVASALSHIAFNALDNDGDEQDNKV--VVKTHNSYALADGEHFIIVDPKLSSTI 685
Db	831	KQLEE-----DNNSNIEWYQNKIEALKKDYESVITSVDKQ--TDIEKLQYKV 876
Qy	686	VDLEK-----ILKLEB-----LDYSTLRQTT--- 707
Db	877	KSLEKEIEBKIRLHYTNVMDETINDDSLRKELEKSKINLTDAYSQIKYKDLVETTSQS 936
Qy	708	--QT-----QFRDMTTAGKALL--SKSNLRQGEKQKQLQEAQFFI--GRVLDLKAIAKAK 757
Db	937	LQQTNSKLDESFKQFTNQIKNLTDERTSLE--DKISLUKEQMFNLNNELDLQKGMEXEK 994
Qy	758	ALVTTKATKNGQLLERSINKAVLAYNNNAIK-----KANVKLEKELDLTLGLVEGKGP 811
Db	995	ADFKKRI-----SILQNNKEVEAVKSEYKSLSKIQNDLQDT----- 1033
Qy	812	LAQATWQGVYLLKTPLEPYEYGLNVYFKSGKLIIVALDMSDTIGE--GQKDAYGNPI 869
Db	1034	-IYANTAQNNY-----EBOLOKHADVSKTISBELRQLHTYKQV 1071
Qy	870	--LNVDED-----NEGYHALAVATLADYEGLDIKTILNSKLSQLTIRSIRQVPTAAVHR 919
Db		
Db	1072	KTMLSRDQLENALKENEKSWSSQKESL--LEQDLDS---NSRIEDLSSQKSL---LYDQ 1123
Qy	920	AGIFQAIQNAAAEABQLLPKPGTHS-----EKSSSSSES 952
Db	1124	IQIYTA---ADKEVNNSTNGPGLNNILITLRRERDILDTKVTVAAERDAKMLRQKISLMDV 1180
Qy	953	ANSKDRGLQSNPKTNRGHRSAIL 975
Db	1181	ELQDARTKLDNSRVEKENHSSII 1203
RESULT 62		
ADS43855		
ID	ADS43855	standard; protein; 1875 AA.
XX	AC	ADS43855;
XX	DT	02-DEC-2004 (first entry)
XX	DE	Bacterial polypeptide #22285.
XX	KW	Recombinant DNA construct; transformed plant; improved plant property;
XX	KW	cold tolerance; heat tolerance; drought tolerance; herbicide; osmois;
XX	KW	pathogen tolerance; pest tolerance; plant disease resistance;
XX	KW	cell cycle pathway modification; plant growth regulator;
XX	KW	homologous recombination; seed oil yield; protein yield; carbohydrate;
XX	KW	nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX	OS	bacterial polypeptide.
XX	FN	Bacteria.
XX	PD	US2003233675-A1.
XX	PD	18-DEC-2003.
XX	PF	20-FEB-2003; 2003US-00369493.
XX	PR	21-FEB-2002; 2002US-0360039P.
XX	PA	(CAOY/) CAO Y.
XX	PA	(HINK/) HINKLE G J.
XX	PA	(SLAT/) SLATER S C.
XX	PA	(CHEN/) CHEN X.
XX	PA	(GOLD/) GOLDMAN B S.
XX	PI	Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX	DR	WPI; 2004-061375/06.
XX	PT	New recombinant DNA construct comprising a promoter positioned to provide
XX	PT	for expression of a polynucleotide encoding a polypeptide from a
XX	PT	microbial source, useful for producing plants with improved properties.
XX	PS	Claim 1; SEQ ID NO 22285; 122pp; English.
XX	CC	The invention relates to a recombinant DNA construct comprising a
XX	CC	promoter functional in a plant cell, where the promoter is positioned to
XX	CC	provide for expression of a polynucleotide encoding a polypeptide from a
XX	CC	microbial source. The invention also relates to a transformed plant
XX	CC	comprising the recombinant DNA construct and a method of producing a
XX	CC	transformed plant having an improved property. The plant is a crop plant
XX	CC	such as maize or soybean. The method of producing a transformed plant
XX	CC	having an improved property comprises transforming a plant with the
XX	CC	recombinant DNA construct and growing the transformed plant, where the
XX	CC	polynucleotide or polypeptide is useful for improving plant properties.
XX	CC	The recombinant DNA construct is useful for producing plants with
XX	CC	improved plant properties, e.g. improved cold, heat or drought tolerance,
XX	CC	tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX	CC	increased resistance to plant disease, better growth rate by modification
XX	CC	of the cell cycle pathway with plant growth regulators, increased rate of
XX	CC	homologous recombination, modified seed oil or protein yield and/or
XX	CC	content, improved yield by modification of carbohydrate, nitrogen or

compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1778 AA;

Query Match 3.8%; Score 195.5; DB 6; Length 1778;
Best Local Similarity 20.2%; Pred. No. 0.01;
Matches 209; Conservative 137; Mismatches 330; Indels 361; Gaps 50;

CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ

44 GADYAESSGSKLKNINETSPPVDDTVTLFSDKRTT--PEKIKD-NLAKGPBEQELKAVT 100
27 GADETTVSEDTAVKTAASATGIESETGSDDETAEPKAEKAEASKETTEKEKAKT 86
101 ENTESEKQI---TSGSQLESKESLSLNKTVPTSTSNWEICDFITKNTLV---GLSKSGV 154
87 EEPASNIKTEINTDKSQLKT---SLKAAVPAGSTY-----NSLFPDNLAK--- 130
155 EKLSDTHLVLPQAAAGTQLIQVASFATPPDKKTAIABYTGRAGENISQLDVGKEI 214
131 -KLAV---IITGNAATGNEVDSAALL-----AISQDL----- 161
215 INGEVFNPSYLLKKVTIPGYKHIGQDAFVNDKNIAEVNLPRESLETISDYAFHAKLQI 274
162 ---SGETGND-----PTDISNIEGLQYLE--NLTSNLSEN--NISDLA-----PLK-- 201
275 DLPDNLKATGELAFPNQITKLSLPRQLMRLAERAFKSNHIKTE--PRGNSLKVIGEA 332
202 ----DLVNLVSLNLSNRVLNLSGVEDLVNQLNVSAN--KALEDISQVASLPVKEI 255
333 SFQDNDLSQMLPDGLEKIESEAFNGPDGHHYNNRVVLWTKSGKNPSGLATENTYVNDP 392
256 SAQGCNKTLEL-----KXPAGAVL----- 275
393 KSLWQSPEDITYKLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIP-----KQHG 444
276 -----PEL-----ETFYQENDLTNLTS--LAKLPKLNLYIKGNASLSKLETLNG 319
445 VT-----ITEIGD-----NAFRNV-DFQNKTLRKYDL 470
320 ATKQLIDASNCTDLETGLDLSGLEMIQLSGCSKLEIYSLKNLPNLVNIATDSAI 379
471 BEVKLPSTIRKIGAFQSN-NLKSFEASDDLBEIK-----EGAFMNN-----RIETL 517
380 EDLGTLLNLPKQLTVLSDNENLTNTAITDLPQLATLTDGCGITSIGTLDNLPKLEKL 439
518 ELKDKLVITGDAAFHINHIYAVLPESVQIEGRSAFRQANNLIIPMG--SKVKTLGEMA 575
440 DLKE-----NQITSI---SEITDLPRLSYLDVSVNNLTITIGLKKLPLEWLN 484
576 FLNRLHLDLSBOKLTETIPQAF---SDNALKEVLLPASLKTIREEAFKKNHLKQLEV 632
485 VSNR-----LSDVSTLTFFPSLNYINNNVIRTVGKMTLPSELKEFAQNNSISDISM 539
633 ASALSHI-----AFNALDNDGDGFQFNKVVVTKHNSYALADGEHPIDVPDKLSSTIV-- 686
540 IHDMPNLRKVDASNLIITIG--TFDN-----LPKLQSLDVHNSRITSTSVIH 585
687 DLEKILKLEGLDYSLRTOTQTFRDMTTAGKALLSKSNLROGEKQKFLQEAQPFPLGRV 746
586 DLPSL-----ETFNAQTNLITNIG--TMDNLPD-----LTYVNLFSNRI 622

747 -----DIDKAIKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKKNVYKLE 796
623 PSLAPIGDLPNLETLIVSDNNSYLSLGTMDG-----VFKLRILDIQ 664
797 KELDILLTGLVEGKGLPAQATWQGVYLLKTLPLPEYVYIGLVNVPDK-SG-----KLIYA 850
665 NNYLNTYCTEGNLSLSDLT-----NLTELNRNNVYIDDISGLSTLSRLIY- 711
851 LDMSTIGEGQKDAYGNPLNVDEDN-RGYHALA-----VATLADYEG 892
712 -----LNDSNKIEDISALSNTLNQELTLENKNIENISALSDLEN 752
893 LDIKTILMSKLSQTSIRQVPTAAHYHRAGIFOA-----IQNAAAEQQLPKP 940
753 LNKLVVSKNIIDISPVANM---VNRGAIVTASQTYTLPTVLSVQSSTIDNPVIWD 808
941 GTHSEKSSSESANSKD 957
809 GTLLAPSSSIGNSGNYKD 825

RESULT 65
ABP40235

ID ABP40235 standard; protein; 3696 AA.

XX AC ABP40235;

DT 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-00134001.

XX 14-AUG-1997; 97US-0055779P.

PR 08-NOV-1997; 97US-0064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

XX N-PSDB; ABN92780.

PT Novel isolated nucleic acid encoding a Staphylococcus epidermis
polypeptide, useful for diagnosing and treating bacterial infections.

PS Disclosure; SEQ ID NO 5080; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
frame (ORF) nucleic acid sequences which encode the amino acid sequences
given in ABP35124 to ABP37960. The *S. epidermidis* sequences have
antibacterial activity and can be used in gene therapy. The sequences can
also be used in the diagnosis and treatment of bacterial infections,
particularly *S. epidermidis* infections. The sequences can be used to
screen for compounds able to interfere with the *S. epidermidis* life cycle
or inhibit *S. epidermidis* infection. N.B. The sequence data for this
patent did not form part of the printed specification, but was obtained
in electronic format directly from the USPTO web site

SQ Sequence 3696 AA;

Query Match 3.8%; Score 195; DB 5; Length 3696;
Best Local Similarity 18.8%; Pred. No. 0.03;
Matches 220; Conservative 164; Mismatches 478; Indels 310; Gaps 45;

QY 10 LTLTVSVVTHNOEVSLVKEPILKOTQASSISGADYAESGKSLKINETSQVDDTV 69
Db 1082 ITHTVNVQKPFQARQALIAKTNEKQASINSNEGTT--IEEKQKAIQSLNDKXNLADQI 1139
QY 70 TDLFSDKR-----TPEKIKONLAK-----GPRQELKAVTENTSEKQ 108
Db 1140 TQASNQVNDALNIGISNISKIOTNFTKQOARDQVQKQOEKAEALNSTPHATQDEKQ 1199
QY 109 TTSQSLQSSESLSINKTVPTSTNWEICDFITKG----- 143
Db 1200 -DALTRLTQAKET-ALNDINOQATQNVDTALTSGIQNTQVNVVRKKQBAKTINDIV 1257
QY 144 -----NTLVGLSKSGVEKLSQTDHLVLPQ-----RADGTLQIV 178
Db 1258 QOHKQSIONDDATTEEKEVANNLVASQOVN--ISKIDNATNTNNDIGIVSDGRQSN- 1314
QY 179 ASFAFTP-----DKKTAIABYTSRAGE-----NGEISOLDVDGKEII- 215
Db 1315 ---ALTPDTSIKENAKNDIDIKAAOKKIKIQINDATDEEQEANRKEIEAKIEAKONIQ 1371
QY 216 -----NEGVFNSYLLKKVTIPTGYKHIGQDAFVNK-----NTAEVNLPSL 258
Db 1372 RNSTRDQVNEAKTNGINKNIENITPATTVKSEARQA-VQNKANEQINHIONTPDATNEBKQ 1430
QY 259 ETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSLPRQLM--RLAERAFKSHI 316
Db 1431 EAINRVS-AELARVOAQNAEHHTQGVKTIKDDAITSLSRINAQVVEKESARNAIEQKAT 1489
QY 317 KTIETPRGSLKVGISAFODNLSQLMLPDGLEKIESBAFTGNPDGDDHYNRNVLTWKSQ 376
Db 1490 QOTQFINNNDATDEKEVANNLVATKQSLDNINSL- -SNDVENVAKVAG 1540
QY 377 KN-----PSGLATENTYVNPDKSLWQSPESPIDYTKWLEEDFTYQKNSVTGFSNKGLOKV 430
Db 1541 INEIANVLPAATVASKAKKIDIDOKLAQIQNIQ-----THQ-TATTEKEAAIOLA 1590
QY 431 KENKN--LEIPKQNGVTITIEIGNAPRNVDFQNKTLRYKLDLEBVKLPSTIRKIGA--- 484
Db 1591 NOKSNEARTAIQNEHSNNGVAQKSNGTHEI-----ELVMPDPAHKHSDAKQS 1637
QY 485 ----FAFOSNNLKSPEASDLEIEIKEGAFMNNRIETLELKDKLVTIGDAAFHINHIAIV 540
Db 1638 INKYNESQNTINT--TPDATDEEKQAL-----DKLTKAKAGY--NKVDQAQ 1682
QY 541 LPEASVOETGRSFRONGANNLIFMGSKVKTGEMAFPLNRLEHLDLSQKQLTEIPVOAQF 600
Db 1683 TNQVSDAKTEAI--DTITNI--QANVAKKPSARVELDSKFEDL-----KRQINATP--- 1730
QY 601 SDNALKEVLLPASLTIREEAPKKNHLKQLEVASALSHIAFNALDDNDGDEQFNKVVK 660
Db 1731 --NATEEE-----KQDAIQRNGKRDEVKNLIN-----QQRDNVEV--- 1764
QY 661 THNGVALADGBHFVDPDKLSSTIVDL-----EKILKILRGL 698
Db 1765 EQHKNIGIQLEETIHANPRKSDALQELQTKFISOTELINNKNKDATNEKEBAKLLIS 1824
QY 699 DYSTLRQTTQO-----FRDMTTAGKALLSKNSLRQGEKQKFLQEAQFPLGR- 745
Db 1825 KXKTTINQAQTNQVONAKONGMNEIATIPATTIKTDAKTDAKAEQOVTTINGNN 1884
QY 746 --VLDKATAK--AEKALVTKATKNGQLLESINKA-----VLAYNN---SAIKKANVR 794
Db 1885 DATDEKAEARKLVKAKIRAKSNITSNTSDEEVNGAKTNGLEKINNIPQSTQTKNAQ 1944
QY 795 -----LEKELDILT-----GLVEGKGPLAQATMVQGVVLLKPLPL 830
Db 1945 EINDKAQEQOLIQINNTPDATTEEKQEAETRNVNAGLAQAIQINNAHSTQEVNESKTNIA 2004
QY 831 PYYITGLNVYFKSKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGVHALAVATLAD- 889
Db 2005 TTKSVQPNVI-----KKEPTAINLQTEANNQKTLIGNDGNATDDEKAAQLVTKQLENEQ 2059

QY 890 ----YEGLDIKTILNSKLSQTSIROVPTAYHAGIFQALONAA--ABAEQLLPKPGTH 943
Db 2060 IQKHESQONQVNVKAQAITAIKLINAHKQDQDAINILTNLAESKKSDIRANQDATT 2119
QY 944 SEKSSSESANSKDRGLQSNPKTKNRGRHSAIL 975
Db 2120 EEKNTAQSID--DTLAQARNNINGANTNALV 2149

RESULT 66
ADS06647
ID ADS06647 standard; protein; 3696 AA.
XX ADS06647;
AC ADS06647;
XX
DT 04-NOV-2004 (first entry)
XX
DE Staphylococcus epidermis polypeptide seqid 5942.
XX
KW antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW recombinant expression vector; infection; computer readable medium;
KW computer based system.
XX
OS Staphylococcus epidermidis.
XX
PN US2004147734-A1.
XX
PD 29-JUL-2004.
XX
PF 01-DEC-2003; 2003US-00724972.
XX
PR 08-NOV-1997; 97US-0064964P.
PR 13-AUG-1998; 98US-00134001.
PR 29-NOV-1999; 99US-00450969.
XX
PA (DOUC/) DOUCETTE-STAMM L.
PA (BUSH/) BUSH D.
XX
XX Doucette-Stamm L, Bush D;
PI WPI; 2004-580138/56.
DR N-PSDB; ADS02875.
XX
XX
XX New isolated polypeptide and encoding nucleic acid derived from
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT treating an S. epidermidis bacterial infection.
XX
PS Claim 17; SEQ ID NO 5942; 741pp; English.
XX
CC The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC given in the specification. Also described are: a recombinant expression
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an S. epidermidis polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for S. epidermidis infection; a recombinant or substantially
CC pure preparation of an S. epidermidis polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection; detecting the presence of a Staphylococcus nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the Staphylococcus genome of
CC commercial importance; a computer based system for identifying fragments
CC of the Staphylococcus plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the Staphylococcus
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the Staphylococcus genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an Staphylococcus epidermidis bacterial

CC infection. This is the amino acid sequence of a S. epidermis protein of
CC the invention.

Sequence 3696 AA:

Query Match 3.8%; Score 195; DB 8; Length 3696;

Best Local Similarity 18.8%; Pred. No. 0.03;
Matches 220; Conservative 164; Mismatches 478; Indels 310; Gaps 45;

Qy 10 L L L T T V S V V T H N Q E V F S L V K E P I L K O T Q A S S I S G A D Y A E S S G S K L I N E T S G P V D D T V 69

Db 1082 I H T H N V N V Q K P O A R Q A L I A K T N E K Q S A I N S D N E G T - - I E E K Q A I Q S L N D A K N L A D R Q I 1139

Qy	70	TDLPSDKR-----	TTPEKIKDNLAK-----	GPREQELKAVTENTSEKQ	108
		:	:	:	
		:	:	:	
Db	1140	TQAASNQVNDNALNIGISNISKIQTNF7KFKQOARDVNVNOKFOSKEAELNSTPHATODESKO	1199		

Qy	109	ITSGSQLEQSKESSLNKTVPSTSNWEICDFITKG-----	143
Db	1200	-DALTRLTOAKET-ALNDINAOACTNONVDLTALTSGIONIO ONTOWNVRKOEAKTTINDIV	1257

Qy	144	-----NTLVGLSKSGVEKLSQTDHLVLPQ-----AADGTQLIQV	178
Db	1258	OOHKOSTONDDATTEKRVANNIWNASQNV--TSKTDNATNTNNGTGVISPGROSTN-	1314

Qy	179	ASFAFTP	-----DKTAIAEYTSRAGE-----	NGEISQLDVGKEII	215
Dh	1315	-----AIIPTPSIKONAKNDIDIRADKTKI	IOBTNDATTFEYICANBTFFATFATKNO	1273	

QY 216 -----NEGEVENSYLLKKVVTPTGYKHIGQDAFVDNK-----NIAEVNLPESL 258

Dh 1372 BNSBDDOINFAKNCINVIENITNTATYKCSBOLVONKAPCINHTONTDADATNEVO 1420

QY 259 ETISDYAPAHALKQIDLPDNLKAIGELAFFDNQITGKLSLPLQM--RLAERAFKSNNHI 316
| | | | |
DB 1431 EAINTVS RELNIVGQVAVHNTCCCTETVDAIITGTCINQVHTCHTANLTPDKAT 1408
| | | | |

Qy	317	KTIEFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEAFNGPGDDHYNRRVVLWTKSG	376
----	-----	---	-----

QY 377 KN-----PSGLATENTVYNPKRSLWQESPEIDYTKWLEEDFTYQKNSVTGFNSKGLQKV 430

QY 431 KRNKN---LEIPKQHNGVITETIGDNARNVDQNKTLRKYDLEEVKLPSTIRKGA--- 484

Qy 485 ---FAFSNNLKSFEASDDLEIKEGAFMNNRIETLEKDKLVTTIGDAAPHNHIYAIV 540
 : ||| : : | : | : | :
 : ||| : : | : | : | :

Qy 541 LPEHSVQEIGRSPFRONGANNLIFWGSKVKTGLGEMAFLSNRLEHLDLSQKQLTEIPVQAF 600

Qy 601 : S0NALKEVLLP\$SLKLTREEAFKKNHLLKOLEVASALS\$HIAFNALDDNDGDEQDFDNKVYVK 660

00	1 7 3 1	---NATIEE---	---NQAALANGANGRADEVANLIN---	QURKRDNEV---	1 7 6 8
Qy	6 6 1	THNSYALADGEHFIVDPKLSITVDL---	---EKILKLEGL	6 9 8	

00	1763	EGRNIGUQEJELIHANFIRKNSDALQKIFISQIETUILLNNKDAINERDEAKRLUEIS	1822
00	699	DYSTLRQTQTQ-----FRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGR-	745

DB	QY	746	--VDLKAIAK--AEKAVTTKKATNGQLERSINKA----
DB		1825	KNKILITLNQAQTNNQVNNAKUNGMEIATIIIPATTIKIDAKVAIDKKAEEQQVTIIINN
DB			1888

DB 1985 DATDEERAEARKLVEKAEI EAKSNIITSDTEREVNGAKTNGLEKINNIQPSTQTKTNAQ 194

QY 795 -----LEKELDLIT-----GLVECKGFLAQATMWQGVLLKTPLEPL 830

DB	1945	EINDRAQSOLIQLNNPTUPATEEEKQEAATNRVNAGLAQAIQINNNAHSTQEVNESKINSIA	2800
QY	831	PEYYIGLVNVPFDSKGLTIYALMDSDTTIGEGQKDAYGNPILNVDEDENGYHALAVATLAD	889

Db	2005	TIKSVQPNVI-----KPETAINSLTQEANNQKTLLIGNDGNATDDDEKEAAKQLVTQKLNEQ
Qy	890	-----YEGLDIKITLNSKLISQLTSIRVPTAAYHRAGIFQAIONAA--AEAQLLLPKPPTH 943

Db	2060	IQIHESHTQDNQVDNVKQAQITAIKLINANAHKRODAINILTNLAESKSDIPANQDATT	2119
QY	944	SEKSSSSSANSKDRGLQSNPKTNRRGHSAIL	975

Db 2120 EEKTAIQSID--DTLAQARNNINGANTALV 2149

RESOLI 67
AAU35908
ID AAU35908 standard; protein; 949 AA.
XX

AC	AAU35908;	
XX		
XX		
DT	14-FEB-2002	(first entry)
XX		
XX		

DE Helicobacter pylori cellular proliferation protein #221.
XX
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design
KW

XX	
XX	
OS	Helicobacter pylori.
XX	
XX	
PN	W0200170955-A2

XX
PD
XX
XX
27-SEP-2001.
21-MAR-2001. 2001W0150009190

XX	
PR	21-MAR-2000; 2000US-0191078P.
PR	23-MAY-2000; 2000US-0206848P.
PR	25-MAY-2000; 2000US-0207722P.

PR 23-OCT-2000; 2000US-0342578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0259311P.
PR 25-DEC-2000; 2000US-0259311P.
PR 25-DEC-2000; 2000US-0259311P.

XX
PA
XX
(ELIT-) ELITRA PHARM INC.

PI Yamamoto RT, Xu HH;
XX
XX
DR WPI; 2001-611495/70.

XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.

PS	Example 3 ; SEQ ID NO 11501; 511pp; English.
XX	The invention relates to antisense inhibitors of genes essential to
CC	The invention relates to antisense inhibitors of genes essential to

CC their use in the discovery of novel antibiotics, the essential genes,
CC prokaryotic cellular processes and the essential genes, and the essential genes,
CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
CC coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
CC and *Proteus mirabilis*.

CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC *Escherichia coli* and *Staphylococcus aureus*. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain

antibodies capable of binding to the expressed proteins are proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of cell lines.

CC of organisms. The present sequence represents an essential prokaryotic

CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 949 AA;

Query Match 3.8%; Score 194; DB 4; Length 949;
Best Local Similarity 19.7%; Pred. No. 0.0052;
Matches 202; Conservative 150; Mismatches 343; Indels 338; Gaps 48;

QY 6 KTVALTLLTVSVVTHNQ--EVFSLVKEPILKQTKTOASSISGADYAESGSKLKINETSG 63
DB 29 KDIGLELKNFMTEPQAGKLYIVDIGKEIQIAN-----OPTKN 69
QY 64 PVDDTVTDLFSKRTTPEKIDNLAGPREQELKAVTENTES-EKQITSGSQLEQSKESL 122
DB 70 PQDNKDDL--NTAATPKPLAKKASKTPKKEBTKAQPCKTKKKEAPAPIIKKEIE 127
QY 123 SLANKTVPSTSNWEICDFITKNTLVGLSKSGVEKLSQTDHLVLPQADGTQLIQVASFA 182
DB 128 IVNTFENQPLVE-----NTPKAVSHSQIEKAKQLQEIQKSGREA-----LNKLQTSN 175
QY 183 FTPDKKTAIAEYTSRAGENGETSOLDVDGKEIINEGEVFNYSYLLKKVVTIPTGYKHIGODA 242
DB 176 TWTNNANSASVNAKK--EISEVKKQEQEIKRH-----ENIKRTGFRVIRKND 224
QY 243 FVDNKNIAFVNLPESLETISDYAFALHALKQIDLPNLKAIGELAFPDNQITGKLSLPRQ 302
DB 225 --ETENETENSVTESKPTQSAAIPEDIKK-----EMQEKQKQETKTKKPSK 271
QY 303 LMLAERAFKSHIYTI-----EPRGNSLVKIGEASFQDNDLSQML-----PDGLSKI 351
DB 272 --PKATPTAKNNKSHKIDFSDVRDPKNDI-----YDDTDEILLFDLHEQDNLNKE 321
QY 352 ESEAFTPGPDHYNRVVLTWTKSGKNPSGLATENTYVPDKSLWQESPEIDYTKWLEE- 410
DB 322 EEE-----KEARQINDRVV---QRNP-----WMNEA 347
QY 411 DFTYQKNSVTGSNKGLOKVRKNKLEIPKQHNGVTITEIGDNAPRV----- 458
DB 348 GIKRQSKKRVPRNDNSQVQISA-IAIPEE--VRVVEFAQKANLADVIKTLFNLGL 403
QY 459 -----DFQNKTLRKVDLSEVKLPSTIRKIGAFQSNLKSFEASDDLEBEKE----- 506
DB 404 MYTKNDFLDKDSIEI LAEFHLEISVQ-----NTLEEFVEVEVLEGVKKRPPVVT 454
QY 507 --GAFMNNRIETLE-LKDKLVLTIGDAAFHINHIYAILVLES-----VQEIGRSAF--R 554
DB 455 INGHVDHGKTSLLDKIRDKRVAHTAAGGITQHIGAYVKEKNKWSPIDTPGHEAFSOMR 514
QY 555 QNGAN-----NLIFM-----GSKVKTLGEM-----AFLSNRLEHLDLSEQKOLTE-- 594
DB 515 NRGAQVTDIAVIVIAAADGVKQOTIEALEHAKAANVPVIFAMNKMWDKENVNPKLKAESA 574
QY 595 -----IPVQAFS-----DNALKEVLLPASLKTIRERAPKKNHILKQLEVASA 635
DB 575 ELGYNPVDMGGEHPIFVSATKGTGDIDNLLLETILIQADIM-----ELKAIIEGSA 624
QY 636 LSHAFNALDDNDGDEQFQDNKVVVTKTHNSYALADGHEFIVD-----PDKLSSTIV 686
DB 625 RAVLVGSEVKEGRG-----AVATVIVSQSTLSVGDSPFAETAPGKVRTMTDDGKSIQ 677
QY 687 DLE-KILKLIBGLDYSTLRQTQTOTFRDMTTAGKALLSKSN-----LRQGEKQKFLQBAQ 740
DB 678 NLKPSMVALITGLS-----EVPVAGSVLIGVENDSIARLQAKREATYLRQ-- 722
QY 741 PFLGRVDLDKATAKAEKA-----LVTKKATKN-----COLLERSINKAVLAYN 783
DB 723 -----KALSKSTKVSFDELSMVAANKELKNIPVIKADTQGLS--BAIKNSLLEIN 771
QY 784 NSAIKKANVGR-----LEKELDLL-----TGLVEGKGPLAQATMVQGVYL 823

DB 772 NEEVAIQIHVSGVGITENDLSLVSSSEHAVILGFNIRPTGNVKNKA----- 818
QY 824 LKTPLPPEYVIGLVNVPYFDKSGKLIYALDMSDTTIGEGQKD---AYGNPILNVD-----ED 875
DB 819 -----KEYNVSIKTY-----TVIYALI-----EGMRSLLGLMSPIEEEHTGQAEV 860
QY 876 NEGYHALAVATLA 888
DB 861 RETFNIPKVGITIA 873

RESULT 68
ADH62806
ID ADH62806 standard; protein; 1182 AA.
AC ADH62806;
XX 15-APR-2004 (first entry)
DT 15-APR-2004 (first entry)
XX Lactobacillus johnsonii mucin binding protein sequence SeqID24.
DE mucin binding; probiotic; glycoprotein; epithelial cell;
KW gastrointestinal tract; lung; uterine cervix.
XX Lactobacillus johnsonii.
XX EPI382970-A1.
FN 21-JAN-2004.
XX 15-JUL-2002; 2002EP-00015609.
XX 15-JUL-2002; 2002EP-00015609.
PR (NEST) SOC PROD NESTLE SA.
XX Pridmore RD;
XX WPI; 2004-111519/12.
DR N-PSDB; ADH62786.
XX Novel mucin binding polypeptide encoded by novel gene of Lactobacillus
PT johnsonii, useful for binding bacteria to mucin.
PS Claim 8; SEQ ID NO 24; 225pp; English.
XX This invention relates to novel mucin binding polypeptides and the DNA
CC sequences which encode them. In particular, the genes are derived from
CC Lactobacillus johnsonii. The invention may allow a better understanding
CC of the mechanism by which probiotics (such as L johnsonii) can bind to
CC mucins, large glycoproteins present on the surface of a large number of
CC epithelial cells including those in the gastrointestinal tract, the lung
CC or the uterine cervix. The present sequence is that of an L johnsonii
CC mucin binding protein of the invention.

XX Sequence 1182 AA;

Query Match 3.8%; Score 194; DB 8; Length 1182;
Best Local Similarity 19.7%; Pred. No. 0.0071;
Matches 222; Conservative 174; Mismatches 421; Indels 308; Gaps 53;

QY 48 AESSGSKLKINETSGVP-DDTVTDLFSKRTTPEKIDNLAGK-----PREQELK 97
DB 84 SSSDGYQKIEV-----PLSDEQVKTLRAGKNDLAIADFNDATNAGTSGTSNKFG 138
QY 98 AVTENTESEKQITSGSQLEQSKESLSLKNKTVPSNSNEICD-----F 139
DB 139 LIIDNNLPQKITVSDGYDMTDDSYTISGYTPKVTYTDKQKEHDLNLSYDEASERF 198
QY 140 ITKGNLTVLGSKSGVEKLSQTDHLVLPQADGTQLIQVASFAFTPKKTAIA---EYTS 196
DB 199 VTKLPVLSVSDYDINVKPYADBEHETLITQKRNIVSLV-----PPKLESLSKVDQRTY 250

Db 1381 -EVASQALAKEKEKALAA-IDQAQTSQVNOAATNGVSAIK-----LIQETKVKPA 1430
 Qy 234 GYKHIGQDAFVDNKNIAEVLNPE-----SLETISDYAPAHAL-----KQID 275
 Db 1431 AREKINOKA---NELRAKINQDEKATABEROVALDKINE--FVNOAMTDITNNRTNQOVD 1485
 Qy 276 LPDNLKATIGELAFPDNQITGKLSLPRQLMLAER-----AFKSNHIKTIEFRGNSLKV 328
 Db 1486 -DTTSQALDSIA-----LVTPDHIVRAAADAVKQOYEAKKREIEQAEHATDEBKQ 1535
 Qy 329 IGEASFQDND-----LSQLMLPDGLEKIESAFTGPDGDDHNNRVVLTWTKSGKNPSGL 382
 Db 1536 VALNQLANNEKRALQNDQAIANNVKKVETNGIATLKG---VQPHIVI-----KPEAQQ 1587
 Qy 383 ATENTYVNPDKSLWQESP-----EIDYTKWL-----EEDFTYQKNSVTGFSN--- 424
 Db 1588 AIKASAENQVESI-KDTPHATVDELDEANQLISDTLKOQOQEIENNTQDAAVTVRNQTI 1646
 Qy 425 KGLQKVK-----RNKN-----LEIPKQHNGVTITEIGD--NAFRNVD 459
 Db 1647 KAIEQIKPKVRKRAALDSIENNNKQDIAIRNTLDTTQDERDVAIDTLNKIVNTIKNDI 1706
 Qy 460 FQNKTLRYDLSE-----VKL-----PSTIRKIGAFAPQSNLKSFEASD-DLBEI 504
 Db 1707 AQNKINAEDVRETETDGNNDNIKVILPKVQVKPAARQSVGVKABQAQAL--IDQSDLSITEE 1764
 Qy 505 KEGA-----FNNRIETLELKDVLITGDAAPHINHIYVLP-----ESVQBIG 549
 Db 1765 RLAAKHLVEQALNQAIQDINHADKTAQVNOQSINAQNIISKIPATTVKATALQIQNIA 1824
 Qy 550 RAEPQNGANN-----LJFMGSVKV--TLGEMAFI-----SNRLEHLD 585
 Db 1825 TNKINLIKANNEATBEONIAIAQVEKELIKAKQOIASAVTNADVAYLLHDEKNEIRIE 1884
 Qy 586 -----LSEQKQLTIPVQAFSDNALKEVLLPASLKTIREAFKKNHLKQLEVASALSHI 639
 Db 1885 PVINRKASAREQLTWL-----FND-----KKQAIENIQATVEE--RNSILAQJ--NIYDT 1932
 Qy 640 APNALDDNDGDEQFON--KVVYKTHNSYALADGSHFIVDPDKLSTTIVDLKILKLIEG 697
 Db 1933 AIGQIDQDRSNAQVDKTASLNLQTIHDL-----DVHPKPKPDPAEKTINDDLARVTALVQ- 1986
 Qy 698 LDYSTLRQTQTPQFDMITAGKALLSKSNLRQGEKQFLQEAQFPLGRVLDLKATAKAEK 757
 Db 1987 -----NYRKVSNNRK-----ADALK 2001
 Qy 758 ALVAKKATNGQLLERSIN--KAVLAYNNSAIKKANVKRLEKELDLTLGLVEGKGLAQ 814
 Db 2002 AITALKQWDEELKTARTNADVDVLKRFNVALSDIEAVITEKNSLLR--IDN---IAQ 2056
 Qy 815 ATMQGVYLLKTPPLPEYIGLVNYVFKSGKLIYALDMSDTIGBGQKXDAGNPIILANVE 874
 Db 2057 QT-A---YAKFKAIAITPELAKVKVLIQD-----YVAD-----GNRR--IDE 2091
 Qy 875 DNEGYHALAVATLADYEGLDIKTLNSKLSQTSIROVPTAAYHRAGIFQALQNAABAE 934
 Db 2092 D-----ATLN-----DIKQHTQFIVDEILAIK-LPAEA-----TKVSPKEI 2126
 Qy 935 QLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG 982
 Db 2127 QPAPKVCPTPIKKEETHSEKVEKE-----LPTQSEG 2158
 RESULT 70
 ID ABM71190
 XX ABM71190 standard; protein; 2186 AA.
 AC ABM71190;
 XX
 XX 20-NOV-2003 (first entry)
 XX Staphylococcus aureus protein #430.
 XX

KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
 KW enzymatic assay; antibiotic target.
 OS Staphylococcus aureus.
 XX WO200294868-A2.
 PN 28-NOV-2002.
 PD 27-MAR-2002; 2002WO-IB002637.
 PF 27-MAR-2001; 2001GB-00007661.
 PR (CHIR-) CHIRON SPA.
 PA Masignani V, Mora M, Scarselli M;
 PI MPI; 2003-120786/11.
 FI N-PSDB; ACF72750.
 DR New Staphylococcus aureus protein, useful as a vaccine for treating or
 DR preventing Staphylococcal infection, specifically an infection caused by
 DR S. aureus, e.g. sepsis.
 PT Claim 1; SEQ ID NO 860; 49pp; English.
 PS
 XX The invention relates to novel genes and encoded proteins from
 CC Staphylococcus aureus. A composition comprising the S. aureus protein, a
 CC nucleic acid encoding the protein, or an antibody to the protein, is
 CC useful as a pharmaceutical, particularly as a vaccine for treating or
 CC preventing infection due to Staphylococcus bacteria, specifically an
 CC infection caused by S. aureus. The composition is particularly useful for
 CC treating or preventing sepsis in a patient. The composition can also be
 CC used for diagnostics. The protein is also used in an assay for enzymatic
 CC studies and as a target for antibiotics. This sequence represents one of
 CC the novel S. aureus proteins of the invention
 XX Sequence 2186 AA;
 SQ
 Query Match 3.8%; Score 194; DB 6; Length 2186;
 Best Local Similarity 19.1%; Pred. No. 0.017;
 Matches 216; Conservative 183; Mismatches 387; Indels 342; Gaps 58;
 Qy 20 HNOEVSLVKEPIIKQTQASSISGADYAESGSGSKLKNETSPGVDVTVDLFSDFD----- 75
 Db 1208 YNAKLAEINATPDATNDEKNAINTLNQDROQAATESIKQANTNAEVDQAATVAENNIDAV 1267
 Qy 76 -----KRTPEKIKDNLKGPQELKAVTENTSEKQITSGSOLESK-----SLSLN 125
 Db 1268 QYDVVKQQAARDKITAQVAK--RIEAVKQTPNATDEBKQ-AAVQINQLKQQAQINQINQN 1324
 Qy 126 KT---VFSTSNWEI-----CDPITKGNTLVGLSGVKEKLSQTDHLVLPQAADGTQL 175
 Db 1325 QTNDQVDTTNNQAVNAIDNVEAEVVIKTAADIEKAVEKQKQIDNSL---DSTDNEK- 1380
 Qy 176 IQVASFAPTPDKTAAIARYTSRAGENGEISQLDVDGKEIINEGEVFNYSLLKVT--IPT 233
 Db 1381 -EVASQALAKEKALAA-IDQAQTSQVNOAATNGVSAIK-----LIQETKVKPA 1430
 Qy 234 GYKHIGQDAFVDNKNIAEVLNPE-----SLETISDYAPAHAL-----KQID 275
 Db 1431 AREKINOKA---NELRAKINQDEKATABEROVALDKINE--FVNOAMTDITNNRTNQOVD 1485
 Qy 276 LPDNLKATIGELAFPDNQITGKLSLPRQLMLAER-----AFKSNHIKTIEFRGNSLKV 328
 Db 1486 -DTTSQALDSIA-----LVTPDHIVRAAADAVKQOYEAKKREIEQAEHATDEBKQ 1535
 Qy 329 IGEASFQDND-----LSQLMLPDGLEKIESAFTGPDGDDHNNRVVLTWTKSGKNPSGL 382
 Db 1536 VALNQLANNEKRALQNDQAIANNVKKVETNGIATLKG---VQPHIVI-----KPEAQQ 1587
 Qy 383 ATENTYVNPDKSLWQESP-----EIDYTKWL-----EEDFTYQKNSVTGFSN--- 424

Db 1588 AAKASAEQVESI-KDTPHATVDELDEANQLISDTLKAQAEIENNTQDAAVTDVRNQT 1646
 QY 425 KGLQKVK-----RNKN-----LEIPKOHNGVTITEGD--NAPRND 459
 Db 1647 KALEQIKPVRRKKAALDSIENKKNQOLDAIRNTLDTTODERDVAIDTLKINVTIKNDI 1706
 QY 460 FQNTLRKYDLBE-----VKL-----PSTIRKIGAFQSNLKSFEASD-DLEBI 504
 Db 1707 AQNKTAEDVDRTEGTGNDNIKVLIPKQVQKPAARQSVGKABQAQNAL--IDQSDLSSTEBE 1764
 QY 505 KEGA-----FWMNRLETLKDLKLVIGDAAPHINHIYALVP-----ESVQETG 549
 Db 1765 RLAAKHLVQALNQAIQDINHADKTAQVNOISINAQNIISKPKPATVVKATALQIQNTA 1824
 QY 550 RSAFRONGANN-----LIFMGSKVK-----TLCEMAFL-----SNRLEHLD 585
 Db 1825 TNKINLIKANNEATDEQNIATAQVEKELIKAKQOIASAVTNADVAIYLLHDEKNEIREIE 1884
 QY 586 -----LSQOKQLTEIPVQAFSDNALKEVLLPASLKTIREAFKKNHLKQLEVASALSHI 639
 Db 1885 PVINRKASAREQLTTL-----FND-----KQQAIEANIQATVEE--RNSILAQLQ---NIYDT 1932
 QY 640 AFNALDDNDGDQFON--KVVVKTTHNSYALADGEHFIIVDPDKLSSTIVDLEKILKLIBG 697
 Db 1933 AIGQIDQDRSNAQVDKTASINLQTIHDL-----DVHPIKPKPAEKTINDDLARVTALVQ- 1986
 QY 698 LDYSTLRQTQTQFRDMTGTAGKALLSKSNLRQEKQKFLQEAQFFLGRVLDLKAIAKAEK 757
 Db 1987 -----NYRKVSNRK-----ADALK 2001
 QY 758 ALVTKATKNGOLLERSIN---KAVLAYNSAIAKKANVRLEKELDLTLGLVEGKPLAQ 814
 Db 2002 AITALKLQMDDELKARTNADVDVLRFNVALSDIEAVITEKENSLLR--IDN---IAQ 2056
 QY 815 ATWQGVYLLKPLPLPEYVIGLVYFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVD 874
 Db 2057 QT-----YAKFRAIATPEQLAKYKVLIDQ-----YVAD-----GNRM--IDE 2091
 QY 875 DNEGYHALAVATLADYEGDLITILNSKLSOLTSIROVPTAAVHRAGIFQAIQNAABAE 934
 Db 2092 D-----ATLN-----DIKQHTQFIVDEILAIAK-LPAEA-----TKVSPKEI 2126
 QY 935 QLLPKPGTHSEKSSSEANSKDRGLQSNPKTNRGRHSAILPRTGSKG 982
 Db 2127 QPAPKVCTPIKKEETHESRKVE-----LNTGSEG 2158
 RESULT 71
 ADH62805
 ID ADH62805 standard; protein; 2209 AA.
 XX AC ADH62805;
 XX DT 15-APR-2004 (first entry)
 XX DE Lactobacillus johnsonii mucin binding protein sequence SeqID23.
 XX KW mucin binding; probiotic; glycoprotein; epithelial cell;
 XX KW gastrointestinal tract; lung; uterine cervix.
 XX OS Lactobacillus johnsonii.
 XX PN EP1382970-A1.
 XX PD 21-JAN-2004.
 XX PF 15-JUL-2002; 2002EP-00015609.
 XX PR 15-JUL-2002; 2002EP-00015609.
 XX PA (NEST) SOC PROD NESTLE SA.
 XX PI Pridmore RD;

XX WPI; 2004-111519/12.
 DR N-PSDB; ADH62785.
 XX Novel mucin binding polypeptide encoded by novel gene of Lactobacillus
 PT johnsonii, useful for binding bacteria to mucin.
 XX Claim 8; SEQ ID NO 23; 225pp; English.
 CC This invention relates to novel mucin binding polypeptides and the DNA
 CC sequences which encode them. In particular, the genes are derived from
 CC Lactobacillus johnsonii. The invention may allow a better understanding
 CC of the mechanism by which probiotics (such as L johnsonii) can bind to
 CC mucins, large glycoproteins present on the surface of a large number of
 CC epithelial cells including those in the gastrointestinal tract, the lung
 CC or the uterine cervix. The present sequence is that of an L johnsonii
 CC mucin binding protein of the invention.
 XX SQ Sequence 2209 AA;
 Query Match 3.8%; Score 194; DB 8; Length 2209;
 Best Local Similarity 19.7%; Pred No. 0.017;
 Matches 222; Conservative 174; Mismatches 421; Indels 308; Gaps 53;
 QY 48 AESSGSKLKINETSGPV-DTPTDLFSDKRTTPEKIKDNLAKG-----PREQELK 97
 Db 1050 SSSDGYQKIEV-----PLSDEQVKTLAGKNDLATAVFDNATNAGTSGTSNKPGEIFG 1104
 QY 98 AVTENTESEKQITSGSQLEBQSKESLSLNKTVPTSTNWEICD-----F 139
 Db 1105 LIIDNNLPQKITTVSDGYDMTDDSYTISGTPEKYVGYTDRKQKEHDLNISYDEASERF 1164
 QY 140 ITKGNLTGLSKSGVKLSQTDHLVLPQQAADGTLQIVASFAFPDPKKTAIA---EYTS 196
 Db 1165 VKLLPLSVSDVDNTNKFYADEHEHTLIYQKRINSLV-----PPKLESUKVDQDY 1216
 QY 197 RAGENGESQLDVGDKIINE-----GEVFNYSYLLK---KVTIPTGYKHIG 239
 Db 1217 TGNBEAKLSQTSSEDTVEVSGKVSDDTDKVAVKVAGTKYSKPTKEHTFAKVPVSYGENT 1276
 QY 240 QDAFVDNKNIAEVNLPESLETISDYAFALHAKQIDLPDLNK-----AIGELAFDQNT 294
 Db 1277 MNVILTDKDGNSSSVKQIVKS--SDRGKTVWSAKDVTDFNGIKFGTTSVNTETENYDPK-T 1334
 QY 295 GKLSIPLROLMLAERAFKSNHIKTIKFGNSLKVIGEASFQNDLSQMLPDLGLEKISE 354
 Db 1335 KGLTLTGKVR-----PTTVIRIGDHTVKKADGTF-----KLVL---DLGRHGAK 1377
 QY 355 APTGNPGD-----DHYNNRWLVMTKSGKNPSGLATENTYVNDKSLWQSPESPIDYTKWLEE 410
 Db 1378 VPPVLIGDVTVDVTQERLTFYVDSN-----NPBLTLNQEKKQSGY----- 1418
 QY 411 DFTYQKNSVTGFSNKGLOKVRKNKMLEIPKOHNGVTITEIGDNAPRNVDFQ--NKLTKYD 469
 Db 1419 -----VPVYTNKEEFKLQTTISDDYPYYSLLINDNNVDAN--WDDVDYNGNKNLKKSF 1469
 QY 470 LEEVKLPSTIRKIGAFQSNLKSFEASDLEEI-----KEGAFMNNRIETLEKLD 521
 Db 1470 SHSVKL-----KEGKNTF---NVVVVDNNDNRSEVQTLVYVYKKAQKIASPOLITATTASD 1521
 QY 522 -KLVTIGDAAPHINHIYAI-----VLPE---SVQETIGRSAFR--QNGANNLI FWSKV 568
 Db 1522 KKSVTVTGKADGNVLYSTDNGNKNVLPEDGVTGVKNGNKLIFKTVKVGNESEVVEYDV 1581
 QY 569 KTLGEMAFPLNRLEHLDSEKQLTEIPVQAFSDNALKEVLLPA-----S 613
 Db 1582 KTIGK-----EESTVDKSVQAQKQD--LRKLDQARALGNTGKYTHESAKK 1625
 QY 614 LKTIREAFK---KN-HLKQLEVAS-----ALSHIAFNALD---DNDGEQDFDNKVVVK 660
 Db 1626 LAQAQFEASKALKDKKNATLQELKQASEQLEEAIKNLVEKFPVDQNKDKEDKOSQNAL 1685
 QY 661 THNSYALADGHEFTVDPDKLSSTIVDLEKILKIEGLDYSTL-----RQTTQTQPRD--- 713

Db 1686 KEKLEETVAGKGF--DKDKYTDDSV-----KVTKALDEAKVVLANKDANSTVDQDAID 1738
 QY 714 -MTTAGKALLSKNSURQEKQKFLQEAQFFLGRVLDKAIA-----KAEKAL---VTK 762
 Db 1739 SIVNATKSLKEK-----QVSPEKTTQEDTPKENKDTSEVLAAGNALKEKADKLSHLDTTK 1794
 QY 763 KATKNGQOLLERSINKAVLAYNNSAIIKANVKELEKELDLITGLVEGKGLAQAT-MVQGV 821
 Db 1795 YTSSEAENLSNALKKVNOVLIN---KDANKAQVQEAOLSL---QAEKDLVEKTESKDV 1848
 QY 822 YLLKPLPLPEYIIGLN-----VYFDKSGKLIYALDMSPTIGEG---OKDAGNPII 870
 Db 1849 ETAKQLREB-----LNKHKDEKSYQYDDSNKV---KDAEKISEGVLDSDKA----- 1894
 QY 871 NVDENEGVALAVA-----TLADYEGLD----- 894
 Db 1895 QADEVNAKADSLVEAEKGLVKEENKPAESDTQEVDDKAREALEQEVKNANVNLIDGYTPE 1954
 QY 895 -----IKTILNSKLSQTSIRQVPTAAVHRAGIFQAIQNA--AEAEQLLPK- 940
 Db 1955 SODKPEILNGVRDVLNDKNASASLEKAEKVLEATATGVLTVQVEHQVELPKVEQPVVTPE 2014
 QY 941 -----GTHSEKSSSESANSKDRGLQSNPKTNRGHRSAILPRTGS 980
 Db 2015 KQTEQEAEKAKSESSESVTSKDEKVEPEEK-KQDAHSVSDKGTGT 2058

RESULT 72

AAU34339
 ID AAU34339 standard; protein; 2434 AA.

AC AAU34339;

DT 14-FEB-2002 (first entry)

DE Staphylococcus aureus cellular proliferation protein #615.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.

OS Staphylococcus aureus.

PN WO200170955-A2.

XX 27-SEP-2001.

PD 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.

PR 26-MAY-2000; 2000US-0207727P.

PR 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.

PR 22-DEC-2000; 2000US-0257931P.

PR 16-FEB-2001; 2001US-0269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS52198.

XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.

XX Example 3; SEQ ID NO 5835; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes

CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2434 AA;

Query Match 3.8%; Score 194; DB 4; Length 2434;
 Best Local Similarity 19.5%; Pred. No. 0.019;
 Matches 226; Conservative 177; Mismatches 454; Indels 300; Gaps 52;
 QY 1 MKKHLKTVALTITTVSVVTHNQEVFSLVKPEILKQTOASSSI-SCADYAESGK----- 53
 Db 226 LKEQVQS-ATTLDGVQTVKNSQTLTAMKGLRDSIANEATIKAGONYTDASPNRNEYD 284
 QY 54 -----SKLKINETSGPV--DDTVDFSDKRTTPEKIKDNLAKGPRE-OELKAVNTES 105
 Db 285 SAVTAAKAIINQTSNTPMEPTITQVTSQVTT-----KEQALNGARNLAQAKTTAKNNLN 339
 QY 106 EKQITSGSQLEQSKESILNKTVFSTNWEICDFITKGNLTVLGSKSVEKLSQTDHL-- 163
 Db 340 NLTSINNAQKDALTRSIDGATTAGV--NQETAKAKATELNNAHSL--QNGINDETQKQTKQ 397
 QY 164 ---VLPQOAGDTQLIQVASPAFTPK-----KTAIAEYTSRAGENGESIQLDVKGKII 215
 Db 398 YLDAEPSKKSAYDAQVNAAKAILTKASQNVDKAAV---EQALQNVNSTKTALNGDAKL 453
 QY 216 NEGEVFNYSLLKQVITPTGYKHI--GQDAFVDNKNIAEVNLPESLETISDYAPHAHLAKQ 273
 Db 454 NEAKAAAKQTLTGLT-----HINNAQRTALDNE-ITQATNVEGVNTVKAKA-----QQ 500
 QY 274 IDLPDLNLKIGELAFFDNOITGK--LSLPQLMLRAERAFKSNHIKTIETFR-----G 323
 Db 501 LD-----GAMQGL-----ETSIRDKOTTLQSQNYQADDAKRTAYSQAVNAATILNKTAGG 552
 QY 324 NSLKVIGEASFQ-----DNDLSQLMLPDGLEKIESEAFTPGPDHNNRVVL---WTKS 375
 Db 553 NTPKADVERAMQAVTQANTALNGIQNLDRAKQAANTAIT--NASDLNTKQKEALKAQVTS 611
 QY 376 GRNPSGLATENT-----YNPDKSLWQESPEIDYTKWLE 409
 Db 612 GRVSAANGVEHTATELNTAMTALKRAIADKAETKASGNYVNADANKRQA-----YD 662
 QY 410 EDFTYQKNSVTG-----FSNKGLO-----KVKRNKNLEIPKQH-----NGVT-- 446
 Db 663 EKVTAENIVSGTPTPTLTPADVTNAATQVTNAKTQLNGHNHLEVAQNANTAIIDLITSL 722
 QY 447 -----ITEIGDNFRNVDVFNQNTLRYDYLEEVLKPS----- 477
 Db 723 NGPQAKLKEQVGQATTLPNVQTVRDNA--QTLNTAMKGLRDSIANEATIKAGONYTDASQ 781
 QY 478 -----TIRK--IGAPAFQSNNLKSPASDDEIKEGAF---MNNRIETLELKDK 522
 Db 782 NKOTDYN SAVTAAKAIIGQTTPSPMNAQEIINQAKQVTAQQAALNGOENLTAQTNAKQH 841
 QY 523 LVTIGD-----AAPHINIYAI VLPESVQEI GRSAFRONGANNLIFMGSKV 568
 Db 842 LNLGSLDLTDAQDKAVKQIEGATHNEV-----TQAQNNADALNTAMNL 886
 QY 569 KT-LGEMAFISNRLEHLDLSEKQKLTIPVQAFSDNALKEVLLPASLKTIREAPK----- 623
 Db 887 KNGIQDQNTIKQGVNFTDADEAKRNAVTVTAQEQILNKAGQNTSKDGVETALENVQR 946

Db 1986 ETELOREHEM--AQTAELQEBL-SGEKNRLAGELQLLEI-----K 2026
Qy 764 ATKNGOLLERSINKAVLAYNNSAIKKA---NVKLEKELDLTGLVEGKGPLAQATMVQ 819
Db 2027 SSXD-QKE-----LTENSELKSLODMHKOQVEGKVRREAEYQLRUHEAKKH 2078
Qy 820 GVYLLKTPLPPEYIYGLNVYFDK-----SGKLIYALDMSDTIG 858
Db 2079 QALLDTON---KQYEVEIQTYREKLTSSKECLSLDILLKSSKEELNNSLKATTOIL 2135
Qy 859 EGOXDA-----YGNPILVNDENEGYHALAVATLADYEGLDIKTILNSKLSOL-----906
Db 2136 BELKTKMDNLKYVNLQKXENERAQGMKLLIKSKQLE--BEKEILOELSOLAQOEK 2193
Qy 907 -----TSIROVPTAAVHRAGIFQIONAAAEQOLPKPG-----THSEKSSSESAN 954
Db 2194 QKTGTWMDTKVDELTEIKE---LKTLEKTKAEDEYLDKYCSLLISHKELEKAKEMLE 2250
Qy 955 SKDRGLQSNPKTNRGHRSAIL 975
Db 2251 TQVAHLCSQOSKQDSRGSPLL 2271

RESULT 75
AAW23996
ID AAW23996 standard; protein; 2482 AA.

AC AAW23996;
XX
XX 28-MAY-1998 (first entry)
DE Human mitosis amino acid sequence.
XX Mitosis; phosphoprotein; mitotic cell cycle; antibody; analogue;
KW inhibition; M phase; Antagonist; hyperproliferative cell; cancer;
KW leukaemia; lymphoma; chromosome segregation.
XX Homo sapiens.

PH Key Location/Qualifiers
FT Domain 258..280
FT /note= "leucine heptad repeat"
FT Domain 340..362
FT Domain 564..593
FT Domain 1387..1443
FT Domain 1885..1962
FT Domain 2146..2188
FT Domain 2165..2187
FT /note= "leucine heptad repeat"
FT Misc-difference 2188
FT Misc-difference 2189
FT Misc-difference 2300
FT /label= Bipartite targeting motif
FT /note= "Optionally C or G"
FT Misc-difference 2301
FT Misc-difference 2303
FT /label= Bipartite targeting motif
FT /note= "Optionally A or T"

XX US5710022-A.

XX 20-JAN-1998.

XX 24-OCT-1994; 94US-00328254.

XX 22-OCT-1993; 93US-00141239.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Lee W, Zhu X;

XX WPI; 1998-109817/10.

DR N-PSDB; AAV09076.
XX New isolated mitosis protein and gene - useful for, e.g. developing
PT products for therapy and diagnosis of hyper-proliferative disorders such
PT as cancers or psoriasis.
XX Claim 1; Col 40-52; 43pp; English.
XX This is the amino acid sequence for mitosis, a phosphoprotein necessary
CC for the cell to enter mitosis. The protein's degradation is also necessary
CC for the cell to advance into the next stages of mitosis. The mitosis
CC protein, can be used to control the growth of cells. An anti-mitosis
CC antibody, a mutant or a non-functional analogue of mitosis can inhibit
CC the mitotic cell cycle by preventing the cells from entering the M phase,
CC and over expression of mitosis or its functional equivalent, would
CC inhibit the cycle by preventing cells from leaving the M phase.
CC Antagonists to this protein can be used to control hyperproliferative
CC cells in, (e.g. thyroid hyperplasia, Grave's disease, psoriasis, benign
CC prostatic hypertrophy, Li-Fraumeni syndrome, breast cancer, sarcomas and
CC other neoplasms, bladder cancer, colon cancer, lung cancer and various
CC leukaemias and lymphomas). Reintroduction or supplementation of lost
CC mitosis function by introduction of the protein or nucleic acid encoding
CC the protein into a cell can restore defective chromosome segregation,
CC which is a marker of progressing malignancy. Malignant proliferation of
CC cells can then be halted. The protein can also be used for the detection
CC and diagnosis of hyperproliferative cells
XX
SQ Sequence 2482 AA;

Query Match 3.8%; Score 192.5; DB 2; Length 2482;
Best Local Similarity 20.2%; Pred. No. 0.025;
Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

Qy 6 KTVALTITVSVVTHNQEVFSLVKEPILKQTOASSISGADYAESGSKLKNINETSQPV 65
Db 1341 KTTALDQLSEKMKETQEL-----ESHQSECLHCQVABAEVKEKTELLQT---L 1387
Qy 66 DDTVTDLFSDKRTTPEKI---KDLAKGPRQELKAVTENTSEKQITSGSQLEQSKS 121
Db 1388 SSDVSELLKDKTHLOEKQLEKDSQALSLTCELENQIAQLNKEKEL-----VKES 1440
Qy 122 LSLNKTVPTSNWEICD-----FITKNTLVGLS-----KSGVEKLS---QTD 161
Db 1441 ESLQARL--SESDYKILNVSKALEALVEKGEPALSLSTQEEVHQLRRCIEKLRVREAD 1499
Qy 162 -----HLV--LPSQADGTQLQVASFAPTPPKKTAIAEYTSRAGENHISOLDVDG--- 211
Db 1500 EKKQLHIAFKLERENDSL-----KDKVENLERELQMSSEENQELVILDAENSKA 1550
Qy 212 -----KEIINGEVFNSYLLKKVTIPTGYKHI-----GQ-----DAF 243
Db 1551 EVETLKTQIEWARSILKVFE---LDVTLRSEKENLTQIQEKQQLSDELKLSFKSL 1607
Qy 244 VDNKNIAEVLNPLETISDYAFALHAKLQIDLPNLKAIGELAPFDNQITGKLSLPQL 303
Db 1608 LEEKEQAEIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 1649
Qy 304 MRLAERAFKSNHIKTIIEFRGNSLKVIGASFDNDLSQLMLPDGLEKTESEFTNGPGD 363
Db 1650 MKATEQSLDPPIIEEHQLR-NSIEKL-RARLEADEKKQLCVLQQLKESEHHA-----D 1700
Qy 364 HYNNEV-----VLWTKSGKNPGLATENTYVNPDKSLWQESPEIDYTKWLEDEFTYQKN 417
Db 1701 LLKGRVENLERELEIARTNQEHAALEAENS-----KGEVTLKAKTEGHTQ--- 1746
Qy 418 SVTGFSNKGLO-----KVKENKNLEIPKQHNGVITIIEIGNAFRNVDFQNKTKRYDL 470
Db 1747 -----SLRGLELDVVTIRSEKENLTNELQKEQERISELEIINSSFENI--LOEKEQEKVQM 1800
Qy 471 BEVKLPSTIRKIGAFAPFOSNNLKSFEASDDLEBEIK-EGAFMNNRIETLELKKLVITGDA 529
Db 1801 KEKS--STAMEMLOTQLKELNERNVAALHNDQACKAKAEQNLSQVECLELEKALQLOGID 1858

Db 2710 QALLDTN---KQYEVETQYREKLTSEKESQKLEIDILLKSSKEELNLSKATTOIL 2766
Qy 859 EGQKDA-----YGNPILNVEDNEGYHALAVATLADYEGLDIKTILNSKLSQL----- 906
Db 2767 BELKTKMDNLKYVNLKKNERAGQKMLIKSCQKLE--EKEILQKLSQLQAQAEK 2824
Qy 907 -----TSIRQVPTAAYHRAGIFQAIQNAABAEQQLPKPG-----THSEKSSSESAN 954
Db 2825 QKTGVMTQKVDLTTTEIKE---LKETLEETKEADEYLDKYCSLLISHEKLEKAKEMLE 2881
Qy 955 SKDRGLQSNPKTNRGHSAIL 975
Db 2882 TQVAHLCSQSQKDSRGSPLL 2902
RESULT 78
ID ADL72172 standard; protein; 3113 AA.
XX ADL72172;
XX
XX
XX 20-MAY-2004 (first entry)
XX Human solid-cancer antigen peptide mitosis.
XX
XX
XX Solid-cancer antigen; FIR; CENP-A; HOOK2; myomegalin; MKRN1; KIAA1545;
XX enigma; TROP2; mitosis; CU-EC-1; cytoetic; solid cancer; human.
XX
XX Homo sapiens.
XX
XX WO2004018518-A1.
XX
XX PD 04-MAR-2004.
XX
XX PF 21-APR-2003; 2003WO-JP005046.
XX
XX PR 23-AUG-2002; 2002JP-00244249.
XX
XX PA (NTSC-) JAPAN SCI & TECHNOLOGY CORP.
XX
XX Shimada H, Hiwasa T, Tomonaga T, Matsushita K, Nomura F;
XX Takiguchi M, Ochiai T;
XX
XX WPI; 2004-248065/23.
XX DR N-PSDB; ADL72171.
XX DR GENBANK; U30872.
XX
XX Human solid cancer antigen peptides and polynucleotides useful in early
XX diagnosis and treating solid cancer.
XX
XX Disclosure; SEQ ID NO 18; 136pp; Japanese.
XX
XX The invention relates to human solid-cancer antigen peptides and encoding
XX polynucleotides. The antigen peptides are selected from FIR, CENP-A,
XX HOOK2, myomegalin, MKRN1, KIAA1545, enigma, TROP2, mitosis, CU-EC-1,
XX Cytostatic; the antigen peptides and their polynucleotides are applicable
XX in early diagnosis and useful in treating solid cancer. With these 10
XX antigen peptides and their encoded polynucleotides, early diagnosis of
XX solid cancer can be achieved accurately and quickly. The present sequence
XX represents a human solid-cancer antigen peptide, mitosis.
XX
XX Sequence 3113 AA;
Query Match 3.8%; Score 192.5; DB 8; Length 3113;
Best Local Similarity 20.2%; Pred. No. 0.034;
Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;
Qy 6 KTVALTTLTVSVVTHNQEVSLVKEPILKQTOASSISGADYAESGSKLKINETSPV 65
Db 1972 KTTALDQLSEKMKERTQSL-----ESHQSECLHCIOVAEAVKEKTELLQT---L 2018
Qy 66 DDTVTLDFDKRTTPEKI---KDLAKGPRQELKAVTENSEKQITSGSLQESKES 121

Db 2019 SSDVSELLKDKTHLOEKLSQKSOALSILTKCELENQIAQLNKEKELL-----VKES 2071
Qy 122 LSLNKTVPSTNWFICD-----FITKNTLVGLS-----KSGVEKLS---QTD 161
Db 2072 ESLQARL--SESDYKLVNKALEAALVEKGEFALRLSSTQBEVHQLRGIEKLRVREAD 2130
Qy 162 -----HLV--LPSQAADGTQLIQVASFAPTPDKKTATAEYTSRAGENGESOLDVDG--- 211
Db 2131 EKKQLHTAELKEREENDSL-----KQVENLERELQWSENOELVILDAENSKA 2181
Qy 212 -----KEIINEGEVFNYSLLKVTIPTGYKHI-----GQ-----DAF 243
Db 2182 EVETLKTQIEBMAKSLKVFZ---LDIVTLSEKENLTQIQEKGQGLSELKLSLSPKSL 2238
Qy 244 VDNKIAEAVNLPESLETISDYAFHAHLAKQIDLDPNLIKAGELAFPNQITGKLSLPRQL 303
Db 2239 LEEKEQAEIQIKESKTAVEM-----LQNLKEL-----NEAVALCGQDEI 2280
Qy 304 MRLAERAPKSNHIKTIEFRGNSLVKIGEASFQDNDLSQMLPDGLEKIESEAPFNGPDD 363
Db 2281 MKATEQSLDPPIEBEHQLR--NSIEKL--RARLEADEKKQLCVLQQLKESEHHA-----D 2331
Qy 364 HYNRV-----VLWTKSGKNPGLATENTVVPDKSLWQESPEIDYTKMLEEDFTYQKN 417
Db 2332 LLKGRVENLERELEIARTNQHALEAENS-----KGEVETLKAKIEGTQ--- 2377
Qy 418 SVTGFSNKGLO-----KVRNKNLEIPKOHNGVTTITEIGDNAPRVDFQNTLRYKYL 470
Db 2378 -----SLRGLELDVVTIRSEKENLTNELQKEQERISELEIINSSFENI--LQEQEKQVOM 2431
Qy 471 BEVKLPSTIRKIGAPAFQSNLKSFEASDDLEIK--EGAPMNNRIETLELKDLVLTGDA 529
Db 2432 KEKS--STAMEMLQTLKELNVAALHNDQEAQKAKEQNLSSQVCELEKALQLOGLD 2489
Qy 530 APHINHIVAIPLPSV---QEIGRSAPRONANNLIFMGSKVKTGLGEMAFSLNRLEHL 585
Db 2490 EAKNY-----IVLQSSVNLQIEV-----EDGKQKL-----EKKDEISLKNQIQ--- 2531
Qy 586 LSEKQKLTIPVQAFSDNAL--KEVLLPASLKTIRREAFKKNHLKQLEVASALSIAFNAL 644
Db 2532 --DQEQLVSKLSQVEGEBHQWKEQNL--ELRNLTVLEQKIQVLQSKNAS-----L 2578
Qy 645 DDNDGDEQFQKVVVTHNSYALADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSTLR 704
Db 2579 QD-----TLEVLQSSYKNLENELELTQMDKMSF-----VEKYNKM-----TAK 2616
Qy 705 QT--TQTOFRDMTTAGKALLSKSNLROGEKOKFLQEAQOFFLGRVLDLKAIAKALVTK 763
Db 2617 ETELQREWHM--AQKTAELQEE--SGEKNRAGELQLLLEEI-----K 2657
Qy 764 ATKNGQLLERSINKAVLAYNNSAIKKA-----NVKRLKELDLTLGLVEGKGPLAQATMVQ 819
Db 2658 SSKD--QLKE-----LTLENSLKKSLDCHMKQDQVEKGVREBIAEYQLRHEAKKH 2709
Qy 820 GVIYLLKPLPPEYIYGLNVYFDK-----SGKLIYALDMSDTIG 858
Db 2710 QALLDTN---KQYEVETQYREKLTSEKESQKLEIDILLKSSKEELNLSKATTOIL 2766
Qy 859 EGQKDA-----YGNPILNVEDNEGYHALAVATLADYEGLDIKTILNSKLSQL----- 906
Db 2767 BELKTKMDNLKYVNLKKNERAGQKMLIKSCQKLE--EKEILQKLSQLQAQAEK 2824
Qy 907 -----TSIRQVPTAAYHRAGIFQAIQNAABAEQQLPKPG-----THSEKSSSESAN 954
Db 2825 QKTGVMTQKVDLTTTEIKE---LKETLEETKEADEYLDKYCSLLISHEKLEKAKEMLE 2881
Qy 955 SKDRGLQSNPKTNRGHSAIL 975
Db 2882 TQVAHLCSQSQKDSRGSPLL 2902
RESULT 79
ADQ18045

FT Domain
 FT /label= Extended_coiled_structure
 FT 280..11350
 FT /label= Extended_coiled_structure
 FT 1380..11610
 FT /label= Globular_domain
 FT /note= "globular_domain consists of 2 direct repeats of
 FT 95 amino acids"
 FT 1620..11750
 FT /label= Extended_coiled_structure
 FT 1850..12990
 FT /label= Extended_coiled_structure
 FT 3048..13248
 FT /label= C-terminal domain
 FT /note= "the C-terminal domain is predicted to form a
 FT proline-rich (10.6%) highly basic (pI 10) globular
 FT domain"
 FT
 PN W09617867-A1.
 XX
 XX 13-JUN-1996.
 PD
 XX 08-DEC-1995; 95WO-US016216.
 PF
 XX 09-DEC-1994; 94US-00353700.
 PR
 XX (FOX-C) FOX CHASE CANCER CENT.
 PA (UYTE-) UNIV TECHNOLOGIES INT INC.
 XX
 PI Yen TJ, Rattner JB;
 XX
 XX WPI; 1996-287116/29.
 DR N-PSDB; AAT34578.
 DR
 XX
 PT DNA encoding kinetochore protein - used as a marker for the G2 and M
 PT phases of a cell cycle, partic. for detection of malignant diseases.
 PT
 XX
 PS Claim 12; Page 41-54; 72pp; English.
 XX
 CC A 372 kDa human kinetochore protein, CENP-F (AAR99795), is detected by
 CC immunofluorescence microscopy only during the G2 and M phases of a cell
 CC cycle. It is the product of a cDNA clone (AAT34578) isolated from a
 CC breast carcinoma cDNA library. Recombinant CENP-F can be produced by
 CC expression in prokaryotic or eukaryotic host cells. CENP-F can be used to
 CC detect autoimmune antibodies to the protein, which may provide an early
 CC diagnosis for the onset of various malignant diseases. Use of CENP-F as a
 CC cell cycle marker allows the specific detection of G2 and M phase cells
 CC
 SQ Sequence 3248 AA;
 Query Match 3.8%; Score 191.5; DB 2; Length 3248;
 Best Local Similarity 20.1%; Pred. No. 0.041;
 Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;
 QY 6 KTVALLTVSVVTHNQVPSLVKPEILKQTOASSSISGADVAESSGSKLKINETSQPV 65
 DB 2059 KTTALDQJSEKKEKTEQEL-----ESHQSECLHCIOVAEAVKEKTELLQT---L 2115
 QY 66 DDTVTLDFSDKRTTPEKI-----KDLAKGPREQELKAVTENTSEKQITSGSLQESKES 121
 DB 2116 SSDVSELLKDKTHLQKLSQLEKDSQALSITKCELENQIAQLNKEKELL-----VKES 2168
 QY 122 LSLNKTVPSTSNWEICD-----FTYKNTLVGLS-----KSGVEKLS---QTD 161
 DB 2169 ESLQARL-SESDYEKLVNSKALEAALVEKGEFALRLSSQTEBHVHQLRRGIEKLRVRIEAD 2227
 QY 162 -----HLV--LPSQADGTLQIQVASFPTPKTKTAIAEYTSRAGENGESOLDVDG--- 211
 DB 2228 EKKQLHIAEKLKERENDSL-----KDKVENLERLOMSENQELVILDENSKA 2278
 QY 212 -----KEINEGEVFNYSLLKKVTIPTQYKHI-----QO-----DAF 243
 DB 2279 EVETLTKQIEEMARSLKIFE---LDLVLRSEKENLTQIQBKQQLSGLDKLLSFFKSL 2335

QY 244 VDNKNIAEVLNLPESLETISDYAPAHALAKQIDIDPNLKAIGELAFPDNQITGKLSLPQL 303
 DB 2336 LEEKEQABIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 2377
 QY 304 MRLAERAFKSNHKTITIFRGNLSKVIGEASQDNDLSQMLPDGLEKTESAFTGNPGDD 363
 DB 2378 MKATEQSLLDPIIEEHQLR-NSIEKL-RARLEADEKKQLCVLQQLKESEHHA-----D 2428
 QY 364 HYNRRV-----VLWTKSGKNPSGLATENTYVVPDKSLMOESPEIDYTKMLBEEDFTYQKN 417
 DB 2429 LLKGRVENLERELEIARTNQSHAALAEANS-----KGEVETLKAKIEGTQ--- 2474
 QY 418 SVTQFSNKGLO-----KVKGNKNLRIPIKOHNGVTTIETGDNAFRNVDFQNKTLRKVDL 470
 DB 2475 -----SURGULEDVVTIRSEKENLTNLOKQEERISELEIINSSFENI-LOEKQOEKVQM 2528
 QY 471 EEVKLPSTIRKIGAFQSFNNLSKFEASDDLEETK-EGAPMNNRIETLEKDKLVITIGDA 529
 DB 2529 KEKS--STAMEMLQTLKELNERVAALHNDQEACAKEQNLSQVCELEKAKQLQGLD 2586
 QY 530 AFHINHIYIVLPESV-----QEIGRSAPROGANNLIFMGSKVTKYTLGEMAFSLNRLEHL 585
 DB 2587 EAKNNY---IVLQSSVKGLIQEV-----EDGKQKL-----EKKDEEISRLKNQIQ--- 2628
 QY 586 LSEQKOLTEIPVOAFSDNAL-KEVLLPASLKTIREBAFKKNHLKQLEVASALSIAFNAL 644
 DB 2629 --DQEQLVSKLSQVGEHQLMKEQN---ELRNLTVELEQKITQVLQSKNAS-----L 2675
 QY 645 DDNDGDEQDNKVVVTHNSYALADGHEHFTVDPDKLSSTIVDLEKILKLTIEGLDYSLR 704
 DB 2676 QD-----TLEVLQSSYKNLENELELTQDKMSF-----VEKNYKM-----TAK 2713
 QY 705 QT-TQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKK 763
 DB 2714 ETELQREMHM--AQKTAELQEEL-SGEKNRLAGELQLLLEEI-----K 2754
 QY 764 ATKNGOLLERSINKAVLAYNNSAITKA-----NVKRELEKELDLLTGLVEKGKPLAQATMVQ 819
 DB 2755 SSKD-QLKE-----LTENSELKKSLDCHMDHQVEKGKVRBEIAYQURLHEAEKKH 2806
 QY 820 GVYLLKTPPLPEYVIGLVVYFDK-----SGKLIYALDMSDTIG 858
 DB 2807 QALLLDTN---KQYEVEIQTREKLTKEECCLSSQKLEIDLLKSKSEELNLSKATYQL 2863
 QY 859 EGQKDA-----YGNPILNVDEDNEGHALAVATLAYEGLDIKTILNSKLSQL----- 906
 DB 2864 BELKKTMDNLKYNQLKENERAQGKMKLLIKSKQLE--EKEKILQKLSQLQAQAEK 2921
 QY 907 -----TSIRQVPTAAYHRAGIFQAIQNAAAAEQQLPKPG-----THSEKSSSESAN 954
 DB 2922 QKTQTVMDTKVDELTTEIKE---LKETLEKTKAEADYLDKYCSLLISHEKLEKAKEMLE 2978
 QY 955 SKDRGLSNPKTNRGRHSAIL 975
 DB 2979 TQVAHLCSQSKQDSRGSPL 2999
 RESULT 81
 ADO84842
 ID ADO84842 standard; protein; 1010 AA.
 XX
 AC ADO84842;
 DT 29-JUL-2004 (first entry)
 XX
 DE E faecalis surface anchored LPXTG protein SeqID3.
 KW LPXTG; cell wall-anchored surface protein; Gram positive bacterium;
 KW extracellular matrix molecule; sequence database; C-terminal;
 KW immunoglobulin-like fold region; Ig-like fold region; antibacterial;
 KW vaccine; gene therapy; infection; medical device; prosthesis;
 KW premature newborn; AIDS; debilitated cancer; bone marrow transplantation.

XX OS Enterococcus faecalis.
XX WO2004025416-A2.
XX PD 25-MAR-2004.
XX PF 15-SEP-2003; 2003WO-US028789.
XX PR 13-SEP-2002; 2002US-0410303P.
XX PA (TEXA) UNIV TEXAS A & M SYSTEM.
XX PA (INH-) INHIBITEX INC.
XX PA (UAB-) UAB RES FOUND.
XX PI Hook M, Xu Y, Sillanpaa JV, Sthanam N, Ponnuraj K, Patti JM;
XX PI Hutchins JT, Hall A;
XX WPI; 2004-315694/39.
XX Identifying LPXTG-containing cell wall-anchored surface proteins from
PT Gram positive bacteria, for treating infection caused by the bacteria,
PT comprises searching sequence information database for the sequence having
PT LPXTG-motif.
XX Claim 16; SEQ ID NO 3; 96pp; English.
XX This invention relates to a novel method of identifying LPXTG-containing
CC cell wall-anchored surface proteins from Gram positive bacteria that bind
CC to an extracellular matrix molecule which comprises searching a database
CC of sequence information for a putative protein sequence having the LPXTG-
CC motif in its C-terminal region and analysing the sequence for the
CC presence of one or more immunoglobulin (Ig)-like fold regions. The
CC invention may be useful for the production of compounds with an
CC antibacterial activity or for production of a vaccine. In addition the
CC disclosed sequences may be useful for gene therapy. The antibody is
CC useful for treating or preventing an infection of Gram-positive bacteria
CC in a human or animal patient. The method and the proteins are useful in
CC generating antibodies for treating and preventing the spread of
CC infections of Gram positive bacteria, for interfering with, or inhibiting
CC binding interactions by Gram positive bacteria, for monitoring the level
CC of gram positive bacterial antigens, or antibodies recognising the
CC antigens in a human or animal patients suspected of containing the
CC devices and prostheses caused by such organisms, and in treating or
CC preventing infections in highly susceptible groups such as premature
CC newborns, AIDS and debilitated cancer patients, and bone marrow
CC transplantation. The present sequence is that of a surface anchored LPXTG
CC protein identified using the method of the invention.
XX SQ Sequence 1010 AA;
Query Match 3.8%; Score 191; DB 8; Length 1010;
Best Local Similarity 19.6%; Pred. No. 0.0087;
Matches 215; Conservative 159; Mismatches 391; Indels 334; Gaps 50;
QY 111 SGSLBQSKESLNNKTPSTSNWEIC--DFITK-----GNTLVGLSKSGVEKLSQTDHL 163
DB 30 SGEKLNQSAKSLAKQSVAKDVQITVKGFINKGTVGNGNTV-----DDQL 77
QY 164 VLPSQAA-----DGTQLIQVASPAFTPKKTAIAEYT 195
DB 78 TIPANVAINEETTPSSLTQWDQVTEATSYEVERDGTVPGNITQNTATFDGSPFISEHT 137
QY 196 SR---AGENG--EISO-----LDVDGKEIINEGEVFNYS-----LLKKVT---IPTGY 235
DB 138 FRVRAVGKNGVSEWSEPIKGTQDDPYKETINQVATSNLPQPGAEKLLKLDKDLSTGW 197
QY 236 KHIGQDAFVDNKNIAEVLNLPESLETISDYAFALALK-----QID-----LPDNLKATG 284
DB 198 -----HTNWSTGANPSDGNF--LSLKFGLGAEYQMDKIEYLPFRNAGN 240
QY 285 EL-----AFFDNQITGKLSLPRQLMRLAERAFKSNHIKTFEPRGN----- 324

Db 241 NILQYRTSKOGANWTFSEPINWKQDALTKTETKQAVRFVEMKVLKSVGNFGSGRE 300
QY 325 -----SLKVIGEASFQDNLSQLMLPGLEKISEAFTG--NPGD----- 362
Db 301 MLFYKQPGTEGILHGDITNDGTIDENDAMSRYNTGLESDVSD--FNGYVERKGDLNKNGVI 359
QY 363 DHYNNRVVLWTKSG-----
Db 360 DAYDISYVLRQLDGGIEIPDVEEIAAGLSLAVVNGKDYLPGLDTLTLFKGQDLKNIN 419
QY 381 GLATENTVNPDKSLWQSPSEIDYTKWLEEDFTYQKNS-----VTGFSNKGLOKVKRNK 434
Db 420 ALSTKMSFDSSKFLVGQPAITNTQOMENYSKYRKHSDNVENLYLVLSNQG-----NK 473
QY 435 NLEIFQKNGVTIITEIGNAFPNVDQNKTLRKYDLEEVKLPSTIRKIGAF----- 485
Db 474 QL-----LNG-----SMDLVTFKVKVKTTRVKRATTVEQPLQDFMSQGLLVG 516
QY 486 -AFQNNLNKSPASDDLEEIKEGAFMNNRIETLELKD-----LVTIGDAAPHINH----- 535
Db 517 QGFQATLSDFSVT-----VKPTELVDKELLQALITLQARVEKEYTPT 561
QY 536 --IYAVLPESVQIEIGRSAPRON-----GANNLI FMGSKVKTLGEMAFLSNRLEHLDLSEQ 589
Db 562 WAIFKPIIDEAVAVLANEQATQTDVSAEAENLEKAAASOLEKMPDVA--NKAD-LEKATQ 617
QY 590 KOLTEIPV--QAFSDNALK--EVLIPASIKTIRBEAFKKNHLKOLEVASALSHPALND 645
Db 618 EGLAKKPSDGEFTBETKVLJESLAAAKQVFAQKVTQEEIDQ----- 661
QY 646 DNDGDEQDNKVVVTHNSVALADGEHFIVDPDKLSSTIVDLEKILKILIEGLDYSTLRQ 705
Db 662 -----ATKTLREAIAQLKEQPVAVNDKETLKEQIAQ--ARGRKPEEGYQFT---K 705
QY 706 TTQTFQDMWTAGKALLSKNSLRQGEKOKFTQEAQOFLGRVDLDKAIKAEKALVTYKKA- 764
Db 706 ETEKQLQBAIQAAEAIVAKETATKEVSEALNALETAMAQLKEVPLVKNKQDLQEVVYKRAQ 765
QY 765 ----TKNQLLERS-----INKAVLAYNNSAIKK--ANVKRLEKELDLTLTCLVEGKGLAQ 815
Db 766 QVTPSEGHQFTASSLQELQKALLAAKNT--LKNPAANQRMIDEAVALTSIDG---LQEE 821
QY 816 TMVQGVYLLKPLPLPEYV---IGLVNFYDKSGKLIYALDMSDTTIGEGOKOAYGNPILNV 872
Db 822 VLVTDKKALEAMIAKAKAIKPSAGKEFTSESKARLTAIDQAEGL-LADKNARQEQI-DI 879
QY 873 DEDNGYHALAVATLADYEGLDIKTLNSKLSQLTSLRQVPTAAVHRAG--IFQAIQNA 930
Db 880 AEKN-----VKTALD--SLE-EQVLQTDKTKLKELLQKAEETLKPAGKQFTKASQ 929
QY 931 AE---AEQLLPKPGTHSEKSSSESANSK-----DRGLOSNPKNRGRHSAILPRTG- 979
Db 930 ABAIKQAKALVEDPNATQEAVDKCLISLSQAIEAMAEBPISNNTGNNGNHSTVSGTGV 989
QY 980 ---SKGSFVYGLVTSVA 995
Db 990 TSQKGATGTTTGTTS 1008
RESULT 82
ABB60349
ID ABB60349 standard; protein; 1216 AA.
XX
AC ABB60349;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7839.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX

OS Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX N-PSDB; ABL04452.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 7839; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1216 AA;

Query Match 3.8%; Score 191; DB 4; Length 1216;
Best Local Similarity 20.1%; Pred. No. 0.011;
Matches 149; Conservative 118; Mismatches 263; Indels 210; Gaps 33;

QY 301 ROLMLABRAPSFKSIKTIIEPRGN--SLKVTIGASFOQNDLSQLMPLDGLKIESEAPT 357
DB 24 QRLHSLRLVAGRNHITLUD--GNDAQHDITURSILHSDNDITEVA-----67

QY 358 GNPQDDHYNRNVLTWTKSGKNPSGLATENTYVNPDKSLWQESP-----EIDY 404
DB 68 --PGGG--SIEVVQNDNSSQPSLAAVQTRLETNSIYPPSPSGDRRTGGRPFQQLQ 123

QY 405 TKWLEEDFTYQKNSVTGFSNKGKQKRNQNL-----EIPQKHNGVITEIGDNAPRV 458
DB 124 LMLDLNRRNIYHVAGNVLPRSLVTMDLSSNLLTVFPQQLPQLPELRIVSLRDLIRSV 183

QY 459 DPQNKTLR--KYDLEVKL--PSTIRKIGAFQSNLKSFEASDLEIEKEGAFNNRIE 515
DB 184 QWKELOVRLRHLERLDGQCINLESDYFQQN-----YSDVHLR 225

QY 516 TLELKDCLVTIGDAAPHINHYIIVLPESVQIGRSFQNGANNLIIFWGSVKVLGMA 575
DB 226 ALNLQSNFVT-----QLPEAV-----PKATGIAHLVAFNAISRHHPSA 264

QY 576 P--LNNRLHLDLSQKQLTEIPVQAFSDNALKEVL-----IPASUKTIRBEAP 622
DB 265 PEGLTETLEYLDL--ERNRLTTVPVALSSHLHLYLTLSNQISQLNLPSPFTENLRVLSL 323

QY 623 KXNHLKQLEVA-----SALSHI--AFNALDD--NDGDEQPD-----NKVVVXTH 662
DB 324 SGNNFMPVLGLKQVTLQSLVMGNSITDPEGIPAVDSWGSNLQTLIRNNKITHLH 383

QY 663 HNSYA-----LADGE-----HFIVDPDKLSSTIVDLKIL 692
DB 384 LGSFAGLEIQIBISLFSNDITTHPLVFNVSRTLKILELSFVFPARSLESLLDPLDALL 443

693 KLIB---GLDYSTLRQTOTQPRDMMTAGKALLSKNLRQGEKQKLOEQAQPLGRVD- 747
DB 444 PLSQLIWGLDNNLKVQSNESFQMBRELSYINLSFNQLKTLPRGLFQSDAHSLVEIDL 503
748 ---LDKAIKAKEKAL-----VTKKATKNGQL-----LERSINKAVLAYNNS 785
DB 504 SYNGLERLEAQTFSHGLDQLTLNLSQNNRLRTIARHAFINLEFLYLDLSYNRLVNSHGA 563
786 AIKKANVKRLEKELDLTLGLVEGKPLAQATMVGVYLLKTPPLPEYIIGLVN----- 839
DB 564 FTVLPNL-----AALDLA-----HNQLCSLSKSLFVVSNTTTP-----RLNVSHNHIA 608
840 -YFDK--SGKLIYALDMSDTIGEGQDAYGN-----PILANVEDNEGVHALAVATLADYE 891
DB 609 SFYDELSSYMYIQDIDISHN-HVTKSDSFTNLANTLRLFLNLAHQNLG--SLQSHAFGDL 665
892 GLDIKTILNSKLSQLTLSIRQ 911
DB 666 FLE---ILNVANNLTSLRR 682

RESULT 83
ADP45530
ID ADP45530 standard; protein; 3616 AA.
XX
AC ADP45530;
XX
DT 26-FEB-2004 (first entry)
XX
DE Chicken AKAP450 amino acid sequence.
XX
KW DISC1 partner; DISC1; ATF4; KIAA1167; KIAA0380; AKAP450; NUDE1;
KW neuroleptic; gene therapy; schizophrenia; psychotic disorder;
KW mood disorder.
XX
OS Gallus gallus.
XX
PN WO2003102587-A1.
XX
PD 11-DEC-2003.
XX
PF 02-JUN-2003; 2003WO-GB002396.
XX
PR 01-JUN-2002; 2002GB-00012856.
XX
PA (UYED-) UNIV EDINBURGH.
XX
PI Millar K, Porteous DJ, Muir WJ, Blackwood D;
XX
DR WPI; 2004-053510/05.
XX
PT Screening for a candidate agent, useful in treating, ameliorating or
PT preventing schizophrenia, psychotic or mood disorder, comprises forming a
PT mixture comprising the candidate agent, a DISC1 partner polypeptide and
PT DISC1 polypeptide.
XX
PS Claim 8; Fig 4G; 89pp; English.
XX
CC The present invention describes a method of screening for a candidate
CC agent (I) which modulates an interaction between a DISC1 partner
CC polypeptide and DISC1 polypeptide, which comprises forming a mixture
CC comprising the candidate agent, a DISC1 partner polypeptide selected from
CC ATF4, KIAA1167, KIAA0380, AKAP450 or NUDE1 or their splice variants,
CC mutant, fragments or orthologues and DISC1 polypeptide. (I) has
CC neuroleptic activity, and can be used in gene therapy. The method is
CC useful for screening a candidate agent, which modulates an interaction
CC between a DISC1 partner polypeptide and DISC1 polypeptide. Mutant forms
CC of ATF4, KIAA1167, KIAA0380, AKAP450 or NUDE1 are useful for identifying
CC agents which substantially restore a wild-type functional activity of a
CC native or mutated DISC1 polypeptide or a mutated form of ATF4, KIAA1167,
CC KIAA0380, AKAP450 or NUDE1. A polynucleotide sequence encoding ATF4,
CC KIAA1167, KIAA0380, AKAP450 or NUDE1 or its fragment, derivative, mutant
CC or orthologue or sequences complementary to the polynucleotide sequences

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3692 AA;

Query Match 3.8%; Score 190.5; DB 6; Length 3692;
Best Local Similarity 19.9%; Pred. No. 0.057;
Matches 230; Conservative 184; Mismatches 399; Indels 343; Gaps 56;

QY 30 EPIKQQTQASSISADYABSSGSKLKINETSFGPDDTVTDLFDKKTTPBEKI--KQNL 87
DB 2696 ENILPATTVKAKADVNAEKEKI-LQINS-----NDEATTEKLVASDNL 2741

QY 88 AKGPREQLKAYTE-----NTESEKQI-----TSGSOLEOS--KESLSL 124
DB 2742 -NHVVETTNQALIEDAPDNNVVEKNKGIGTTRDLOPLVVKPTAKSIESAVEKKTEI 2800

QY 125 NKTVPSTNWEICDFTTKNTLVLSKSGVEKLSQTHLVLPSSAADGTQLIQVAFPT 184
DB 2801 NQTONATHD-EVREGNLQNLHHEKAKDNVQ-SQTNQV---ENAEQNSLDQINN--FR 2853

QY 185 PD---KKTALAYTSRAGEN--GEISQ-----LDVQGEIIN-----E 217
DB 2854 PDFSKKRNABIV-KAQQNKIDEIEQEFSAEQEBKONALQHLDEQVKEIINSINQAND 2912

QY 218 GFVFN-----YLLKVTIPTGYKHIGQDAFVNDKNIAEVN-LPESELETIS 262
DB 2913 NEVDNAKTSGLNITEYRPNYKKNALIKLYD-----VSDTQEAINGVPDATE--D 2963

QY 263 DYAFHLALKQIDLPDNLKAIGELAFDQNTG-----KLSLPR-----QJM 304
DB 2964 ELQEANSKLNKI-LDDAKQIG-LAHTNNEVDIYNEVSQKMTILPRVDTKAVARSVLN 3021

QY 305 RLAEAPFKS-----NHK-----TIEFRGSLKVIAGEASFQDNLS 340
DB 3022 ALAKQLIKTFENTADVTHEERDANHNKVEQLSLVFNAIEKDKRIQVADQLFGLNELN 3081

QY 341 QLMLPDGLEKIESEAPTPGDDHNNRVLVTKSGKNPSGLATE---NTYVNDPKSLWQ 397
DB 3082 SIFINITQKTPARKAISGWA--SQLNNSI-----NNTPYATEERQIALNKVKAIVD 3131

QY 398 ESPP-IDYTKWLEEDFTYQNSVT-----GF-----SNKGLQKVRKNKLEI 438
DB 3132 DANEKIREANTDSEVLGTGKSNATILLOAISADVQVQKPAFEINAAQBIQERINGSDA 3191

QY 439 PKHNGVITEIG---DNAFRNVDFQNTLRKYDLEE--VKLPSTIRKIGAFQFQSNL 493
DB 3192 TREKKEALKQVDTLVNHSFIFINNKNQEVYDQDKTIEAHKIPKISTIKPOLNEI 3251

QY 494 SPEASDDLEEKEGAFMNNRIETLEK---DKLVITIG-----DAAPHINHIAVL 542
DB 3252 TIQLDTQDRLIK-----NKESTVEKASIDKLKTAARIAESIDKA-QTNEEVNKK 3305

QY 543 ESQVQIGR---SAFRQNGANNLIFMGSKV-----KTIGEMAFILSNRLE----- 582
DB 3306 QSIDISKILPVIEKSAARNEIHOKEAIEVIRGLINDNEEATKEEDKIALNQLDITLTQAN 3365

QY 583 -HLD-----LSEQKQLTEIPVQAFSDNALKEVLLPASLKTIREAPKK-NHLKQLEVA 633
DB 3366 VSIDQALNWEANRAKSTANSEINKISVIAIKK---PEAIAEIQBLADKCKLKPQSEA 3422

QY 634 S-ALSHIAPNALDDGDGDFQFNKVVVTKHNSYALADGEHIVDPDKLSSTIVDLKIL 692
DB 3423 TIEEQSALNELEQ-----ALKSAINHIHOSQNNESVSAALKESI 3462

QY 693 KLIEGLDYSTLRQTOTQFRDMMTAGKALLSKNLRQGEKQFKQEAFFLGRVDLDAI 752
DB 3463 SLIDSIEIQAHKLEAKAYIDGSDDKINDISSRATNEEKQIFVSKLALINRT--HKQI 3520

QY 753 AKAEKALVTXKATKNGQLLERSINKAVLAYNNSAIKKANV-KRLEKELDLTLGLVEGKGP 811
DB 3521 DEASTFVSVEIVRNFVKEADKLNSIVR-----KKAKASKEIELEADHVKQMINAN-- 3571
QY 812 LAQATMVQGVYLLKTLPLPLPEYIYGLNVYFDKSGKLIYALDMSDTIGEGQKDAYGNPILN 871
DB 3572 LSASTRVK-----QNTARLINEIVSNALSO-----LN 3598
QY 872 VDENEGHALAVLATLADYEGLDIKTILNSKLSQTSITRQVPTAAHYHRAGIFQAIQAAAA 931
DB 3599 KVTNNK-----EVDEIVNETIEKLKSIQ-----IR 3623
QY 932 EAEOLLKPGCTHSEKSSSESANSKDRGLQSNPKTNEGRHSAILPRTGSKGSFVYIGLY 991
DB 3624 EDKILSQRSTSMTEKSNQCYSENNTIKSLPEA--GNADKSLP-----LAGV 3670
QY 992 TSVALLSLITAIKKKK 1007
DB 3671 TLISGLAIMSRKKKK 3686

RESULT 85

ADN22762

ID ADN22762 standard; protein; 925 AA.

AC ADN22762;

XX 02-DEC-2004 (first entry)

DE Bacterial polypeptide #5415.

XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.

OS Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 5415; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the

the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2398 AA;

Query Match 3.7%; Score 189; DB 6; Length 2398;
Best Local Similarity 18.7%; Pred. No. 0.039;
Matches 210; Conservative 153; Mismatches 444; Indels 314; Gaps 46;

QY 2 KKHKTVALTLTVSVVTHN-----QEVFSLVKEPIKQTOASSISGDAVSESGSKS 55
DB 1111 KQHLG-----TLTHITTAQRDLTNQISOATNLGAVESVKQN--ANSLDGA-----MGNLQ 1159
QY 56 LKINETSGBV-----DDTVTDLFSDKRTTPEKID-----NLAKGPREQELKAV---- 99
DB 1160 TAINDKSGTLASQNFLEDADEQKRNAYNAQVSAETILNKQTGPNTAKTAVEQALNNVNA 1219
QY 100 -----TENTESEKO-----ITSGSOLEQ-SKESL--SLNKTVPTSTSNWEICDFITKNT 145
DB 1220 KHALNGTQNLNNAKQAATAINGASDLNQKQDALKAKAQAQVRSNAQDVQHNATLNT 1279
QY 146 LVGLSKSGV-----EKLSDTHLV-----LPSQAADGT 173
DB 1280 ANGLKHAADKNTLASSKYVNAUSTKQNAITKVTNAEHLISGTPVTVPVSEVTAA 1339
QY 174 QLIQVASFAPTP-----KKTAAIAYTSRAGENGESIQOLDVD 210
DB 1340 NOVNSAKQELNGDERLREAKQONATAIDALTQLTNPQAKLKEQVQANRLQEDVQTVTN 1399
QY 211 GKSEINEGEVFNLSLLKKVTIPTGYKHIGQDAFVDNKN-----IAEVNLPESL 258
DB 1400 GQALNNAMKGLRDSANETTVKTSQNY--TDASPNQSTYNSAVSNKGIINQTNPN--- 1454
QY 259 ETISDYAFAPHAHLKQIDLPDLNKAIGELAFQNOITGKLSL-----PRLMLRAERAF 311
DB 1455 -TWDTSAITQATOVNNAKNGLAENLR--NAQTAKQNLNTLSHLTNQKSAISSQID 1511
QY 312 KSNHKTIFRGNLSKVIAGEASFQNDLSQMLPDLGLEKIESEAF--GNPGDDHYNRRV 369
DB 1512 RAGHVSEVATKNAAT--ELNTQMGNLQAIHQDQTVK-QSVKFTDADKAKRDATYNAV 1567
QY 370 ----VLWTKSGNPSGLATENTYNNPKLSLWQESPEIDYTKWLBEDFTYQKNSVTGSN 424
DB 1568 SRAEAILNKTOGANTSQKQVEAAIQNVSSA-----KNALNGDQ 1606
QY 425 KGLQKVRKNLEIPQKHNGVTITEIGDNAFRNVDPQNTLRKYDLEEVKLPSTKIKGA 484
DB 1607 -----VTNAKN-----AAKVALNLTSSINNAQRDLTKIDQATTVAGVEA 1647
QY 485 FAFQSNLKSFFASDDLEEBKEGAFNNRIETLEK-----DKLVTIGDAFPHNIHY 537
DB 1648 VENTSTQLTAMAN-----LQNG--INDKNTLASENYHDADSKKTAVTQAVTNAENIL 1700

QY 538 -----AIVLPESVOEIGRSEAFRQNCANNLIIFWGSVKVTKLGEMAFLSNLEHLDLSE 588
DB 1701 NKNSGSLDKTAVENALSOVANAAGALNGHNHLEQAKGNANT-----TINGLQHLTTAAQ 1754
QY 589 QKQLTEIPVQAFSDNALKEVLLPASL-----KTIREEAFKKNHLKQLEVASALS 639
DB 1755 KDKLQKQVQQAQNVAGVDTVKSSANTLNGAMGTGLRNSIQDNTATKNGQNYLDATERNKTN 1814
QY 640 AFNALDDNDG-----DEQFNKVVVVKTHNSYALADGEHFIVDPDKLSSTIVD-- 687
DB 1815 YNNAVDSANGVINATSNPNMDANINQIATQVTTSTKNAL-DGTHNLTKAQKATATNAIDGA 1873
QY 688 --LEKILK-----LIEGLD-----YSTL 703
DB 1874 TNLNKAQKDALKAQVTSQORVANVTISIQQTANELMTANGLOHQHGIDDENATQTKYRDA 1933
QY 704 RQTTQTPRDMTMTAGKALLSK---SNLRQGEKQKFLQEA---OFFLGRVDLDKAIKAE 756
DB 1934 EQSKKTAYDQAAVAAKAILNKQTSNSDKAAVDRLAQVTVSTPKDALNGDKAKLAEAKAAK 1993
QY 757 KALVTKKATKQGL--LERSINKAVLAYNNSAIKANVKRLEKELDLTLGLVEGKGLAQ 814
DB 1994 QNLGTLNHTNAQRTDLEGOINQATTVDGVNTV-KTNANTLIDGAMNLSQGSINDK----D 2048
QY 815 ATMVQGVYLLKTPPLPEYIYGLNVVFDKSGKLIYALDMSDTIGEGQXDAGNPIILNVDE 874
DB 2049 APLRNQNYL-----DADESKRNATQVATAEGILNKQTGGN-TSKADV 2091
QY 875 DNEGHALAVATLAD--YEGLDIKTILNLSKLSQTSIRQVP 913
DB 2092 DN-----ALNAVTRAKAALNGAD--NLNRAKTSATNTIDGLP 2126
RESULT 88
ABP29526
ID ABP29526 standard; protein; 1240 AA.
XX AC ABP29526;
XX DT 02-JUL-2002 (first entry)
XX Streptococcus polypeptide SEQ ID NO 8228.
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX Streptococcus pyogenes.
XX WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;
PI Tettelin H;
XX WPI; 2002-352536/38.
XX DR N-PSDB; AEN70157.
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX

PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voos H;
 XX
 DR WPI; 2002-010914/01.
 XX
 PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
 PT and prevention of *Listeria* and related bacterial infections, and related
 PT polypeptides.
 XX
 XX Claim 6; SEQ ID NO 456; 192pp; French.
 XX
 CC The present invention relates to the genome sequence of *Listeria*
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in *L.*
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of *L. monocytogenes* and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by *L.*
 CC monocytogenes and related organisms. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2013 AA;

Query Match 3.7%; Score 188.5; DB 5; Length 2013;
 Best Local Similarity 21.9%; Pred. NO. 0.032;
 Matches 250; Conservative 148; Mismatches 429; Indels 315; Gaps 66;

QY 18 VTHQEVSLVPEILKQTOASSISGADYASSGSKLKNETSGPVDDT--VTDL---- 72
 DB 975 IITLNEAIVPTDAPRDEIAYNSIAYRIDKETGSGTSLASEPPRGVSKTNPSTDLNTA 1034
 QY 73 ---FSDKRTTPEKIKDNLAGPREQELKAV-----TENTESEK---QITSGSQLEOSKE 120
 DB 1035 GNSFSD-----LNKNGVKDADELGNVAVKLDLYKKNNEPEKVEVTYTSSDALDKGK 1087
 QY 121 SLSLN-----KTPVSTSNWEICDFITKG-NTLVGLSKSG-----VEKLSQTDHLVLP 168
 DB 1088 LFDENGLNNGTYKIAAHLFN-KNADFITTPGNKIVKDSKDDSIGMTVNNSTFTI--DD 1144
 QY 169 AADGTQL-----IQVASFATPP-----DKKTAIAYTSRAGENGISQLDVGDKREI 214
 DB 1145 LANGPNKLVIGIQPIYMATPIKGAFFVFNKDGEPKITSY---GDYBIALFDKDGKEV 1200
 QY 215 -----INRGV-ENSYLLK-----KVTIPGYKHIGQDAFVNDKNIAEVLNPLESLET 260
 DB 1201 QSAIKTNSKGFESFNDVAIKNPADFKLVKTAPTGTNFTV-----YSAKN----- 1243
 QY 261 ISDYAFAPHALAKQIDLPDNLKAIGELAFFDQNTIGKLSLPRQLMRLAERAFKSNHIKTIE 320
 DB 1244 ----PLFNNMSTKEYTLNSVVPVGGVA--EIIYT-ETSKPTTKI-ILDKAVTPNAI-TIE 1294
 QY 321 FRGNSLKVIGBASFOQNDLSQLMPLDGLKIESEAPTGN-----PGDDHNNRNVVLWTSKG 376
 DB 1295 SSDEATEVTNWTLESDSGTVVY-----TGTGNTIRIPNDE-----GTYYA 1335
 QY 377 KNPSGLATNTVNPDKSLWQESPEIDY---TKWLEEDFTYQKNSVTFPSNKGLOKVKRN 433
 DB 1336 KNTATDEAGNT-ASDSEKTP-----DIDYTVPTLTWNQDASAEVNSTE---ANIGWIK---P 1384
 QY 434 KNLEIPKQHNG-----VTITEIGDNA-----FRNVDPQNKI- 464

DB 1385 LNVAAATDTHDGNITPVDYDYSKVMDVLGTYPVTVTATDASGNKATQTVNLRIVDTTSPTI 1444
 QY 465 -----LRKYDLEEVKLPSTIRKIGA-----FAFQSNLKLK 494
 DB 1445 LITNPLIYSTENMRK--LTQELYTAAGLIGDNDYDLAPQAOVQPNQKQPMWFTSNFSTI 1502
 QY 495 FEASDDLEIEKEGAFM-----NNRIE---TLELKDKLVTICDAAAFHINHIYAIYL 541
 DB 1503 F---GDIASVRAGOVQVQVNLADSSGNOAIPQTIINVVD---TVGPVKADNVSHVNT 1556
 QY 542 PESVQIEGRSAPRQNGANNLIPMGSKVKTLSGMAFLSNRLEHLDLSE--QKOLTEIPVOA 599
 DB 1557 TKTEAEFPQDARLDVTDNN-----DDTTDLIITSNFAEKVNLNKGKYEVTITATDT 1608
 QY 600 PSDNALKEVLLPASL-KTIREAPKKNHLKQLEVASA--LSHIAFNALDDNDGDQFQDNK 656
 DB 1609 KGNQTKETITVQVSKDKPVTADPKISYQKIEVTEANFLSGVHAETDELGDGVKITSD 1668
 QY 657 VVVKTHHN---SYALA---DGEHFIVDPDKLSSTIVDLKILKILIEGLDYSLRQTQT 709
 DB 1669 FAEKVDFNKGVTYTYTLNAKDEYGNTPAEFKVSVSI--FNKIAPTFNADNKTIEAVNEL 1726
 QY 710 QFRD--MTTAGKALLSKNLQGEKQKFLQEAQFFLGRVLDLDAIAKAEKALVTYKATK- 766
 DB 1727 PSLESIFPKIEAKDYLSGNKL---KVTYTPK-QTIKGNVPGEYSI-----KVTTKD 1772
 QY 767 -NGQLLESINKAVLAYNNSAIKKA-NVKRLE---KELDLL-----TGLVEG---KGP 811
 DB 1773 DSGNIAETVTLTKDTGTPSKIMTKSKLEVOSEKPNWITFFGIKATDIVDGDVTKNI 1832
 QY 812 LAQATMVQ---GVYLLKTLPLPEYIYICLVNVPYFDKSGKLIYALDMSDTIGSGQDAYGN 867
 DB 1833 KVDSEVNLNKVGTY-----PIYFTVTDALGNESKLSKSTVQIVDT-----SS 1875
 QY 868 PILNVEDNEGYHALAVATLADYEG-LDIKT-ILNS--KLSQTSIRQVPTAAYHRAGIF 923
 DB 1876 PELATIDKLEISYP--TGKTISDKQLQDIGTKVTNSYGVTKVTTNLSKI--VDMNKAGKY 1931
 QY 924 QAIQNA-----AAAEQOLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGHSAIPLR 977
 DB 1932 KVTVTATNSSGGVAEKITILL-----TVKNSDSSFIAPVSKD-----DNKNKPAKNIPK 1979
 QY 978 TG 979
 DB 1980 TG 1981
 RESULT 90
 ABU32684
 ID ABU32684 standard; protein; 2013 AA.
 XX
 AC ABU32684;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by *Prokaryotic essential gene #18211*.
 XX
 KW Antisense; *prokaryotic essential gene*; cell proliferation; drug design.
 OS *Listeria monocytogenes*.
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykend JW;
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI: 2003-029926/02.
 XX N-PSDB; ACA36554.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 XX PT for homologous nucleic acids required for cellular proliferation to
 XX PT isolate candidate molecules for rational drug discovery programs.
 XX
 XX Claim 25; SEQ ID NO 60608; 1766pp; English.
 XX
 XX The invention relates to an isolated nucleic acid comprising any one of
 XX the 6213 antisense sequences given in the specification where expression
 XX of the nucleic acid inhibits proliferation of a cell. Also included are:
 XX (1) a vector comprising a promoter operably linked to the nucleic acid
 XX encoding a polypeptide whose expression is inhibited by the antisense
 XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
 XX polypeptide or its fragment whose expression is inhibited by the
 XX antisense nucleic acid; (4) an antibody capable of specifically binding
 XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 XX proliferation or the activity of a gene in an operon required for
 XX proliferation; (7) identifying a compound that influences the activity of
 XX the gene product or that has an activity against a biological pathway
 XX required for proliferation, or that inhibits cellular proliferation; (8)
 XX identifying a gene required for cellular proliferation or the biological
 XX pathway in which a proliferation-required gene or its gene product lies
 XX or a gene on which the test compound that inhibits proliferation of an
 XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
 XX compound's activity; (11) a culture comprising strains in which the gene
 XX product is overexpressed or underexpressed; (12) determining the extent
 XX to which each of the strains is present in a culture or collection of
 XX strains; or (13) identifying the target of a compound that inhibits the
 XX proliferation of an organism. The antisense nucleic acids are useful for
 XX identifying proteins or screening for homologous nucleic acids required
 XX for cellular proliferation to isolate candidate molecules for rational
 XX drug discovery programs, or for screening homologous nucleic acids
 XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 XX the target prokaryotic essential genes. Note: The sequence data for this
 XX patent did not form part of the printed specification, but was obtained
 XX in electronic format directly from Wipo at
 XX ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 2013 AA;
 XX
 XX Query Match 3.7%; Score 188.5; DB 6; Length 2013;
 XX Best Local Similarity 21.9%; Pred. No. 0.032;
 XX Matches 250; Conservative 148; Mismatches 429; Indels 315; Gaps 66;
 XX
 XX 18 VTHNQEVSFLVKEPIKOTQASSISIGADYABSSGSKLKNETSGPVDDT--VTDL--- 72
 XX 975 IYLFNEAIVPTDAPNDEIAYNSIAYRIDKGTSGTSLKASEPPRGVKSNTSPSDDLIA 1034
 XX 73 ----\$GDKRTTPKIKDNLAGPREQLKAV-----TENTESEK---QITSGSQLEQSK 120
 XX 1035 GNSFSD-----LNKNGVKDADELGLNAVKLDLYKONNEPEKVTVTYSSDALDK 1087
 XX 121 SLSLN-----KTVPSISNWEICDPITKG-NLYVLGSKG-----VEKLSQTDHLVLP 168
 XX 1088 LFDNGLNNGTVTKIAAHLN-KNADFITTPGNKIVKDSKDDSIGMTVNNSTFTI--DD 1144
 XX 169 AADGTQL-----IQVASFATP-----DKTAIABYTSRAGENGESQLDDVGKEI 214
 XX 1145 LANGNPKNLVGLQIPIYMATPIKGAUVFNKDGEPKITSY-----GDYEIALFDKQKEV 1200
 XX 215 -----INEGV-FNSYLLK-----KVTIPTGYKHIGQDAFVDNKNIAEVLNLPESLET 260
 XX 1201 QSAINTNSKGEPSFNDVAIKNADPKLVKTAFTGTNFV-----YSKKN----- 1243

QY 261 ISDYAPAHALAKQIDLPNLKAIGELAFFNQITGKLSLPQLMRLAERAFKSNHIKIE 320
 DB 1244 ----PLFNMSTPKYTLANSVPGVGVA--EYIIT-ETSKPTTKI-ILDKAVTPNAL-TIE 1294
 QY 321 PRGNSLKVIGASQDNDLSQLMLPDGLEKTESAFTGN---PGDDHYNNRVNVLWTKSG 376
 DB 1295 SSDEATEVTNWTLESDSGTVVY-----TGIGNTIRIPNDE-----GTVIA 1335
 QY 377 KNPGLATENTYVNPDKSLMOESPEIDY---TKWLEBEDFTYKNSVTVGFSNKGLOKVKRN 433
 DB 1336 KNTATDEAGNT-ASDEKTF-----DIDYVPTLVNQDASAEVNSTE--ANIGWIK---P 1384
 QY 434 KNLSEIPKQNG-----VTITEIGNA-----FENVDFQNK- 464
 DB 1385 LNVAATDTHGNIITPVVDYSKVKWDVLGTVPTVTTATDASGNKATQTVNLRIVDTTSPTI 1444
 QY 465 -----LRKYDLEBVKLPSTIRKIGA-----PAPOSNNLKS 494
 DB 1445 LITWNPITYSTENNRK-LTEQELYTAAGLIGDNDYDLAPQAVQPNKQPMVFTSNFSTI 1502
 QY 495 PEASDDLEEIKEGAPM-----NNRIE---TLELKDVLVTIGDAAFHINIYAILV 541
 DB 1503 F---GDIASVKAGQYQVQVNLADSSGNAIPQTITINVD---TVGPVIKADNVSYHVNT 1556
 QY 542 PESVQEIGRSFRQNGANNLIFMGSKVKTGEMAFNLNLEHLDLSE--OKQLTEIPVQA 599
 DB 1557 TKTEAEFPQDARLOVTNN-----DDTDLIITSNFAEKVNLNPKGKYEVTITATDT 1608
 QY 600 FSDNALKEVLLPASL-KTIREAPFKQNLKOLEVASA--LSHIAFNALDDNDGDEQFONK 656
 DB 1609 KGNQTTKEITVQVSKDPVITADPKISVQKIEVTEANFLSGVHAEVTDGLDGVKITS 1668
 QY 657 VVVKTHN---SYALA---DGEHPIDDPKLSSTIVDLKILKLEGLDYSTLRQTQT 709
 DB 1669 PAEKVDFENKVGTYTTLNAKDEYGTAEAPVKVSVSI--FNKIAPTFFNADNKTIEAVNEL 1726
 QY 710 QFRD--MTTACKALLSKSNLQSGKQKFLQEAQFPLGRVLDLDAKAEKALVTKKATK- 766
 DB 1727 PLSISIFKIEAKOYLSGNKL-----KVITYTPE-QTIKGNVPGEYSI-----KVTKD 1772
 QY 767 -NGQLLERSINKAVLAYNNSAIKKA-NVKRLE---KELIDL------TGLVEG---KGP 811
 DB 1773 DSGNIAETVTLTIKDTTPGSIKMTSTKLEQVSKEPNWTTPFGIKATDIVDGDVTKNI 1832
 QY 812 LAQATWQ---GYVLLKTPPLPEYIIGLVNVPYDKSGKLIYALDMSDTIGEGQKDAYGN 867
 DB 1833 KVDSEVNLNKVGTY-----PIYFTVTDALGNESSEKLSKSTVQIVDT-----SS 1875
 QY 868 PILNVDENEGYHALAVATLADYBGL-DIKT-ILNS--KLSQLSIRQVPTAAVHRAGIF 923
 DB 1876 PELTIDKLEISYP--TGKTISDKQFLQDIGTKVTSYGVTKVTNNLSKI--VDMNKAGKY 1931
 QY 924 QAIONA-----AAEAEQLPKPGTHSEKSSSSSSSANSKORGLQSNPKTNRGHSAIILP 977
 DB 1932 KVTYATNSSGGVAEKITILL---TVKNSDSSFTAVPSKO-----DNKNKPAKNIPK 1979
 QY 978 TG 979
 DB 1980 TG 1981
 RESULT 91
 ADA89551
 ID ADA89551 standard; protein; 1448 AA.
 XX
 XX ADA89551;
 XX
 XX DT 20-NOV-2003 (first entry)
 XX
 XX Staphylococcus aureus antigenic protein #90.
 XX
 XX antigenic protein; vaccine; Staphylococcus aureus; pathogenic organism;
 XX antibacterial; neuroprotective; immunosuppressive; antiinflammatory;

KW antiulcer; immunostimulant; ophthalmological; pathogenic microbe;
 KW bacteraemia; septic shock; organ infection; skin infection;
 KW bacterial basal colonisation; bacterial eye infection; septicaemia;
 KW tuberculosis; food poisoning; blood infection; peritonitis; endocarditis;
 KW sepsis; meningitis; pneumonia; stomach ulcer; gonorrhoea; toxic shock;
 KW necrotising fasciitis; impetigo; histoplasmosis; Lyme disease;
 KW gastro-enteritis; dysentery; shigellosis; skin disorder.
 XX Staphylococcus aureus.

OS WO2003011899-A2.

PN 13-FEB-2003.

XX 02-AUG-2002; 2002WO-GB003606.

XX 02-AUG-2001; 2001GB-00018825.

PR 09-JAN-2002; 2002GB-00000349.

XX (UYSH-) UNIV SHEFFIELD.

PA (BIOS-) BIOSYNEXUS INC.

XX Foster S, Mond J, Clarke S, McDowell P, Brummel K;

PI WPI; 2003-256434/25.

DR New antigenic polypeptides from Staphylococcus aureus or S. epidermidis,

PT useful as a vaccine for immunizing humans against e.g. bacteraemia, septic
 PT shock, septicaemia, tuberculosis, meningitis, pneumonia, gonorrhoea or
 PT impetigo.

XX Claim 4; Page 140-141; 189pp; English.

XX The present invention describes an antigenic protein or its part, which
 CC is for use as a vaccine. The antigenic protein is encoded by an isolated
 CC DNA molecule of Staphylococcus aureus or S. epidermidis partial gene
 CC sequences (designated dnaSA and dna SE, respectively; and which encodes a
 CC protein expressed by a pathogenic organism. Also described: (1) a vaccine
 CC composition comprising at least one antigenic protein; (2) a method of
 CC immunising an animal against a disease or condition caused by a
 CC pathogenic microbe by administering the antigenic protein or the vaccine;
 CC (3) an antibody or its binding part obtainable by the method above; (4)
 CC preparing a hybridoma cell line producing monoclonal antibodies; (5) a
 CC hybridoma cell line produced by the method of (4); and (6) identifying
 CC opsonic antigens expressed by a pathogenic microbe. The antigenic
 CC proteins have antibacterial, neuroprotective, immunosuppressive,
 CC antiinflammatory, antifungal, immunostimulant and ophthalmological
 CC activities, and can be used in vaccines. The antigenic proteins or
 CC vaccines can be used for immunising an animal (specifically a human)
 CC against a disease or condition caused by a pathogenic microbe, e.g.
 CC bacteraemia, septic shock, organ infection, skin infection, bacterial
 CC basal colonisation, bacterial eye infections, septicaemia, tuberculosis,
 CC bacteraemia-associated food poisoning, blood infections, peritonitis,
 CC endocarditis, sepsis, meningitis, pneumonia, stomach ulcers, gonorrhoea,
 CC strep throat, streptococcal-associated toxic shock, necrotising
 CC fasciitis, impetigo, histoplasmosis, Lyme disease, gastro-enteritis,
 CC dysentery, shigellosis, S. aureus-associated septicaemia, food-poisoning,
 CC skin disorders, S. epidermidis-associated septicaemia, peritonitis or
 CC endocarditis. The present sequence represents a S. aureus antigenic
 CC protein sequence from the present invention.

XX Sequence 1448 AA;

Query Match 3.7%; Score 187.5; DB 6; Length 1448;

Best Local Similarity 19.3%; Pred. No. 0.024;

Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

QY 12 LTTVSVVTHNQEVFSLVKEPILKQTCASSISGA-----DYAESGSKLXINETSGPV 65

DB 140 ITAEETAAANADVNAV-----TQANSNIEAANSQNDVDQAKTTGNSI-----183

QY 66 DDTVTDLFSDKRTTPKPKIDNLAQPREQELKAVTENTESEKQITSGSQLEQSKESLSLN 125

Db 184 -DQVPTVNMKKATARNEITAIL--NNKLQEIQATPDATDEEKQ-----AADAENANGKAN 237
 QY 126 KTVPTSNWEICDFITKGNLTVGLSKSGVKEKLSQTDHLVLPQAADGTQLIQVAFPTP 185
 Db 238 QAIASATT-----NAQVDEAKANAE-----AANAVTP 265
 QY 186 D--KKTATAEYTSRAGENGESQLDVGKEIINEGEVFNFSYLLKKVTIPTGYKHGQDAF 243
 Db 266 KVVVKQAQAKD-----EIDQLQATQTNVNN-----DQDAT 295
 QY 244 VDNKNIAEVLNPLESETISDYAFALHALKQIDLPNLKAIGELAFDFNOITKLSLPRQL 303
 Db 296 TEER-----EAAIQQLATAVTDKKNITA-----ATDDNGV-----326
 QY 304 MRLAERAFKSNHIKTIEFRGNSLKVIGEASFQNDLSQMLMPLDGLKEKIESEAFTPNGD 363
 Db 327 -----DQAKDAGKNSIOSTQPATAVKSNKNDVDQAVTTQNOAID 366
 QY 364 HYNRRVVLWTKSGKPSGLATENTYVNPDKSLWQSEPIDYTKWLEEDFTYQKNSVTGFS 423
 Db 367 -----NTTGATTEER--NAAKDLVLKAKEKAYQDILNAQTT---NDVTQIK 407
 QY 424 NKGLOKVRNKNLEIPKOHNGV-----TITEIGDNFAFRNVDFQNKTL--RKYDLEEVKLS 477
 Db 408 DQAVADIQ-----GITADTTIKDVAKDELATKANQKALIAQTADATTEERQ 455
 QY 478 TIRKIGAFAPQSN-NLKSFEASDDLEETKEGAFM-----NNRIETL-ELKDK 522
 Db 456 ANQQVDAQLTQGNQNIENAAQSIDDVNTAKDNAIAIDPIQASTDVKTNRABELLTEMONK 515
 QY 523 LVTIGDAAFPHNHYAIVLPESVQSIG--RSAPFRONGANNLIFMGSKVKVTLGEMAFLSNR 580
 Db 516 ITEILLNNETNE-----EKGNDIGPVRAAY-EEGLNNI---NAATTG-----555
 QY 581 LEHDLBSQKQJTEIPVQAQFSDNALKEVLLPASLTIREEAF-KKNHLKQL-----630
 Db 556 ----DVTTAKOTAVQVQQLHANPVKK---PAGKKELDQAAADKKTQIEQTPNASQOEIN 608
 QY 631 ----EVASALSHIAFNALDDNDGDGFQFNKV-----VVKT--HHSNYALADGHP--674
 Db 609 DAKQEVDTLNOAKTN--VDQSTNTEYVDNAVKEGKAKINAVKTFSEYKKDALAKIEDAYN 667
 QY 675 ---IVDPD---KLSTIIVDL-EKILKJEGLDYSTLRQTTQ-----TQFRDMT- 715
 Db 668 AKVNEADNSASTSSEIAEAKQKLAELKQADONVQATSKDDIEVQIHNDLNDINDYTI 727
 QY 716 TAGKA-----LLSKSNLRQGEKQKFL---QEAQFFLGRV-----D 747
 Db 728 PTGKKESATTDLYAYADQKKNISADTNTATQDEKQQAIKQVDQNVQTALESINNGVDNGD 787
 QY 748 LDKATAKAEKALVTKKATNGQOLLERSINKAVLAYNNSAIKANVUKLEKELDILTG---804
 Db 788 VDDALTQGAAL---DAIQVDAVTVPKANQAI-----EVKAEDTKESIDQSDQJTAEBK 838
 QY 805 -----LVEGKGPLAQATMVQGVYLLKTLPLPEYIYGL---NVYFKSGK---846
 Db 839 TEALAMIKQITDQAKOGITDATTAEVKAQAQ-----GLEAFDNIQIDSTEKQKAI 890
 QY 847 ---LIYALDM-----SDTTGEGQKDAYGNPILNVDENEGYPHALAVATLADY-EGLDI 895
 Db 891 BELETALDQIEAGVNVNADATTE-EKEAFTNAL-----EDILSKATEDISDQTTNAEI 942
 QY 896 KTLINLSKLSLTISIRQVETAYVHRAGI-----FOAIONAAAE-----934
 Db 943 ATVKNSALEQLKAQRIINPEVKKNALAEIREVVNKQIEIKKNADADASAKETARTDLGRYF 1002
 QY 935 ----QLLPKPGTHSEKS-----SSSESANSKDRGLQSNPKTNRGRHSAIL 975
 Db 1003 DRPADLKTQTNAEVASELQNTVPIAEIAIVPQNDPDANDTNGDNDNATANSANATP 1062
 QY 976 PRGTG 979
 Db 1063 ENTG 1066

QY 976 PRTG 979
 Db 2093 ENTG 2096
 RESULT 94
 ABJ19002
 ID ABJ19002 standard; protein; 2478 AA.
 XX AC
 XX ABJ19002;
 XX 06-MAR-2003 (first entry)
 XX Pathogen specific antigen related staphylococcal protein SEQ ID No 220.
 XX Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
 KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
 KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
 KW autoimmune disease; HIV; hepatitis.
 XX Staphylococcus sp.
 OS WO200259148-A2.
 PN 01-AUG-2002.
 PD 21-JAN-2002; 2002WO-EP000546.
 PF 26-JAN-2001; 2001AT-00000130.
 PR (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 PA Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;
 PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;
 PI Tempelmaier B;
 XX MPI; 2003-075410/07.
 XX Identifying, isolating and producing hyperimmune serum-reactive antigens
 PT from a pathogen, for preparing vaccine or medicament for treating or
 PT preventing e.g. staphylococcal infections, comprises providing antibody
 PT preparation.
 XX Example 7; Page 186; 252pp; English.
 XX The invention relates to a novel method for identifying, isolating and
 CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
 CC allergen, a tissue or host prone to auto-immunity, where the antigens are
 CC used in a vaccine, comprises providing antibody preparation from a plasma
 CC pool of a type of animal, or individual sera with antibodies against the
 CC specific pathogen, tumour, allergen, tissue or host prone to auto-
 CC immunity. The hyperimmune serum-reactive antigens comprising any of the
 CC 62 sequences of 53-2261 amino acids fully defined in the specification,
 CC or their hyperimmune fragments are useful for the manufacture of a
 CC pharmaceutical preparation, particularly a vaccine against staphylococcal
 CC infections or colonisation against *S. aureus* or *S. epidermidis*. The
 CC preparation of antibodies is useful for the manufacture of a medicament
 CC for treating or preventing staphylococcal infections or colonisation
 CC against *S. aureus* or *S. epidermidis*. The antibody preparations may also
 CC be used for diagnostic and imaging purposes. Other conditions that can be
 CC treated include cancer, autoimmune diseases or infections caused by viral
 CC (e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This
 CC sequence represents a staphylococcal protein relating to the method for
 CC identifying and producing pathogen specific antigens of the invention
 XX Sequence 2478 AA;

Query Match 3.7%; Score 187.5; DB 6; Length 2478;
 Best Local Similarity 19.3%; Pred. No. 0.05;

Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

12 LTTVSVVTHNQEVFLVKPEILKQTQASSISGA-----DYAESSGSKLKINETSGPV 65

Db 1170 ITAEIEIAAADVDNAV-----TQANSIEAANSQNDVDQAKTTGENSI----- 1213
 QY 66 DDTVTDLFSKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLESLSLN 125
 Db 1214 -DQVPTVNNKATARNEITAIL--NNKIQEIQTATPDATDEKQ---AADAENANGKAN 1267
 QY 126 KTVPSTNWEICDFITKGNLTVLGSKGVEKLSOTDHLVLPQADGTQLIOVASFATP 185
 Db 1268 QAIQAATT-----NAQVDEAKANAE-----AAINAVTP 1295
 QY 186 D--KKTALAEYTSRAGENGESQLDDVQKEIINEGEVFNLSYLLKKVTIPTGYKHIGQDAF 243
 Db 1296 KVVKKQAQK-----EIDQLQATQTNVIN-----DQNA 1325
 QY 244 VDNKIAEAVNLPLESLETISDYAFALHALKQIDLPNKLKAGELAPFQNDQITGKLSRQL 303
 Db 1326 TEEK-----EAAIQQLATAVTDKNNITA---ATDDNGV----- 1356
 QY 304 MRLAERAPKSNHIKTIIEPRGNSLVIGESAFQDNDLSQLMLPDGLEKIESAFTGNPGDD 363
 Db 1357 -----DQAKDAGKNSIQSTQPATAVKSNAKNDVDQAVTTQNOAID 1396
 QY 364 HYNRRVVLWTKGKGNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFS 423
 Db 1397 -----NTTGATTEEK--NAAKDLVLKAKEKAYQDILNAQTT---NDVTQIK 1437
 QY 424 NKGLQVKRNKNLEIPKQHNGV-----TITEIGDNFRNVDFQNTL--RKYDLEEVKLPS 477
 Db 1438 DQAVADIQ-----GITADTTIKDVAKDELATKANEQKALIAQTADATTEBEQ 1485
 QY 478 TIRKIGAPAFQSN-NLKSPEASDDLEIKEGAFM-----NNRIETL-ELKDK 522
 Db 1486 ANQOVDAQLTQGNQNIENAOISIDVNTAKDNAIQADIPQASTDVKTNNARALELTEMQNK 1545
 QY 523 LVTIGDAAFHINHIYAVLPESVOEIG--RSAPFGNANNLIWFGSKVKTLEGMAFLSNR 580
 Db 1546 ITEILNNNETNE-----EKNDIGPVRAAY-EEGLNNI-----NAATTG----- 1585
 QY 581 LEHLDLSEKQLTETIPVQAFSDNALKEVLLPASLKTIREEAF-KKNHLKQL----- 630
 Db 1586 ---DVTAKDTAVQKVQQLHANPVKK---PAGKELDQAAADKKTKTQIEQTPNASQBEIN 1638
 QY 631 ----EVASALSHIAFNALDDNDGDEQFDNKV-----VVKT---HNSYALAGEHP-- 674
 Db 1639 DAQOEVDTLNOAKTN-VDQSTNEYVDNAVKEGKAKINAVKTFSEYKCKDALAKIEDAYN 1697
 QY 675 --IVDDP-----KLSSTIVDL-EKILKLEGLDYETLQTTQ-----TQFRDMT- 715
 Db 1698 AKVNEADNSNASTSEIAEAKQKLAELKQTADQNVNQAATSKDDIEVQHINDLDNINDYTI 1757
 QY 716 TAGKA-----LLSKSNLRQGEKQKFL-----QBAQFPLGRV-----D 747
 Db 1758 PTGKESATTDLYAVADQKKNISADTNTATQDEKQAIKQVDQNVQTALESINNGVDNGD 1817
 QY 748 LDKAIKAEKALVTKATKQGLLERSINKAVLAYNSAIKKANVKLEKELDLTG--- 804
 Db 1818 VDDALTQGAIAI---DAIQVDAVTKPKANQAI-----EVKAEDTKESIDQSDQLTAEEK 1868
 QY 805 -----LVEKGKGLAQTVMQGVVLLKTPLPPEYIGL-----NVYFDKSGK--- 846
 Db 1869 TEALAMIKQITDQAKQIGITATTTAEVEKAKAQ-----GLEAFDNIQIDSTEKQKAI 1920
 QY 847 --LIYALDM-----SDTIGEQKDAYGNPILNVDNEDNEGYHALAVATLADY-EGLDI 895
 Db 1921 EELESTALDQIEAGVNVNADATTE-EKEAFTNAL-----EDILSKATEDISDQTTNAEI 1972
 QY 896 KTIILNSKLSOLTSIROVPTAAYHRAGI-----FQAIQNAARAE----- 934
 Db 1973 ATVKNLSALQKQAKRINPEVKKNALAIREVNVNQIIEIKNADADASAKETARTDLGRYF 2032
 QY 935 ---QLLPKPTHSEKS-----SSSESANSKDRGLQSNPKTNRGRHSAIL 975

Db 2033 DRFADKLDKQTNAEVAELQNVTTIPAEIAIVPQNDPDANDTNGINDNDATANSANATP 2092

Qy 976 PRTG 979

Db 2093 ENTG 2096

RESULT 95

ABM71899

ID ABM71899 standard; protein; 2478 AA.

AC ABM71899;

XX 20-NOV-2003 (first entry)

XX Staphylococcus aureus protein #1139.

XX Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;

XX enzymatic assay; antibiotic target.

XX Staphylococcus aureus.

XX WO200294868-A2.

XX 28-NOV-2002.

XX 27-MAR-2002; 2002WO-IB002637.

XX 27-MAR-2001; 2001GB-00007661.

XX (CHIR-) CHIRON SPA.

XX Masignani V, Mora M, Scarselli M;

XX WPI: 2003-120786/11.

DR N-PSDB; ACF73459.

XX New Staphylococcus aureus protein, useful as a vaccine for treating or
PT preventing Staphylococcal infection, specifically an infection caused by
PT S. aureus, e.g. sepsis.

XX Claim 1; SEQ ID NO 2278; 49pp; English.

XX The invention relates to novel genes and encoded proteins from
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a
CC nucleic acid encoding the protein, or an antibody to the protein, is
CC useful as a pharmaceutical, particularly as a vaccine for treating or
CC preventing infection due to Staphylococcus bacteria, specifically an
CC infection caused by S. aureus. The composition is particularly useful for
CC treating or preventing sepsis in a patient. The composition can also be
CC used for diagnostics. The protein is also used in an assay for enzymatic
CC studies and as a target for antibiotics. This sequence represents one of
CC the novel S. aureus proteins of the invention

XX Sequence 2478 AA;

Query Match 3.7%; Score 187.5; DB 6; Length 2478;

Best Local Similarity 19.3%; Pred. No. 0.05; Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

Qy 12 LTVSVVTHQEVFLVKEPILKQTAQSSISGA-----DYAESGSKLKNETSGPV 65

Db 1170 ITAEIIAANADVDAV-----TQANSNIEAANSQNDVDQAKTTGENSI----- 1213

Qy 66 DDTVTDLDFDKTTEKIKONLAKPREQELKAVTENTSEKQITSGSOLEKSKLSLN 125

Db 1214 -DQVFTVNNKATARNETAIL--NNKLOEIQTATPDATDEEQ---AADAEEANTENGKAN 1267

Qy 126 KTVPTSNWEICDFITKNTLVGLSKSGVEKLSQTDHLVLPQADGTQLIQVAFPTP 185

Db 1268 QAISAATT-----NAQVDEAKANAE-----AANAVTP 1295

Qy 186 D--KKTATAEYTSRAGENGESQLDVGDKGEIINEGEVFNYSLLKKVTIPTGYKHGQDAF 243

Db 1296 KVVKKQAAKD-----EIDQLQATQTNVNN-----DONAT 1325

Qy 244 VDNKNIAEVLNPESETISDYAFALHALKQIDLPNLKAIGELAFDFNOITKLSIPLROL 303

Db 1326 TEEK-----EAAIQOLATAVTDAKNNTA-----ATDNGV----- 1356

Qy 304 MRLAERAFKSNHIKTIIEFRGNSLKVIGBASFOQNDLSQLMLPDGLEKIESEAFNGPDG 363

Db 1357 -----DQAKDAGKNSIQSTQPATAVKSNAKNDVDQAVTTQQAID 1396

Qy 364 HYNRRVLTWTKSGKNPSGLATNTVNPDKSLWQESPEIDYTKYKLEEDTYOKNSVTGFS 423

Db 1397 -----NTTGATTEEK--NAAKDILVKAKEKAYQDILNAQTT--NDVTQIK 1437

Qy 424 NKGLQKVKRNKNLEIPKQHGV---TITEIGDNAPRVNDFQNKTL--RKYDLEEVKLP 477

Db 1438 DQAVADIO-----GITADTTIKDVADELATKANEQKALIAQTADATTEKEQ 1485

Qy 478 TIRKIGAFAPQSN-NLKSFEASDDLEIEKEGAFM-----NNRIETL-ELKDK 522

Db 1486 ANQVDAQLTQGNQNIENASQIDDVNTAKDNAIQADPIQASTDVKTNRABELLTEMQNK 1545

Qy 523 LVTIGDAAAFHINHIYIVLPESVQIG--RSAFRONGANNLIFMGSKVKTLGEMAFLSNR 580

Db 1546 ITEILNNNETTNE-----EKGNDIGPVRAAY-EEGLNNI--NAATTTG----- 1585

Qy 581 LEHLDLSEKQKLTETIPVQAFSDNALKEVLLPASLTIREEAF-KKNHLKQL----- 630

Db 1586 ----DVTTAKDTAVQVQOLHANPVKK---PAGKKELDQAAADKKTKQTEQTPNASQOEIN 1638

Qy 631 ----EVSALSHIAFNALDDNDGDQFDNKV-----VVKT--HNSYALADGEHF-- 674

Db 1639 DAKQEVDTLQAKTN-VDQSTNEYVDNAVKEGKAKINAVKTFSEYKQDALAKLEDAYN 1697

Qy 675 --IVDPD-----KLSSTIVDL-EKILKLEGLDYSTLROTTQ-----TQFRDMT- 715

Db 1698 AKVNEADNSASTSSEIAEAKQKLAELKQTADQNVNQATSKDDIEVQIHNDLDNINDYTI 1757

Qy 716 TAGKA-----LLSKSNLRQEKQKFL---QEAQFFLGRV-----D 747

Db 1758 PTGKESATDLYAYADQKKNISADTNTATQDEKQQAIKQVDQNVQTALESINNNGVDNG 1817

Qy 748 LDKATAKAEKALVTKKATKNGOLLERSINKAVLAYNNSAIKKANVKRLEKELDLTG--- 804

Db 1818 VDDALTQCKAAL--DAIQVATVVKPANAQI-----EVKAEDTKESIDQSDQITAEK 1868

Qy 805 -----LVGEGKPLAQATMVQGVYLLKTPPLPHYIYL-----NVPFKSGK---- 846

Db 1869 TEALAMIKQITDQAKQIGITDATTAEVEKAKAQ-----GLEAFDNIQIDSTEKQKAI 1920

Qy 847 --LIYALDM-----SDTIGEGQKDAYGNPILNVDEDEGHHALAVATLADY-EGLDI 895

Db 1921 EELETALDQIEAGVNVNADATTE-EKEAFTNAL-----EDILSKATEDISDQTTNABI 1972

Qy 896 KTIILNSKLSQTSIROVPTAAVYHRAGI-----FOAIONAAAEAE----- 934

Db 1973 ATVKNASALEQKQRIINPEVKNNALEAIREVVNQKIEIKKADADASAKETARTOLGRYP 2032

Qy 935 ----QLLPKPGTHSEKS-----SSSESANSKDRGLQSNPKTNRGRHSAIL 975

Db 2033 DRFADKLDKQTNAEVAELQNVTTIPAEIAIVPQNDPDANDTNGINDNDATANSANATP 2092

Qy 976 PRTG 979

Db 2093 ENTG 2096

RESULT 96
ABU07438
ID ABU07438 standard; protein; 3210 AA.
XX
AC ABU07438;

XX 28-JAN-2003 (first entry)
 XX Protein differentially regulated in prostate cancer #41.
 DE Prostate cancer; gene expression; differential regulation;
 KW molecular marker; drug target; cancer detection; cancer diagnosis;
 KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
 XX Homo sapiens.
 OS
 XX WO200281638-A2.
 PN
 XX 17-OCT-2002.
 PD
 XX 08-APR-2002; 2002MO-US010824.
 PF
 XX 06-APR-2001; 2001US-0281731P.
 PR
 XX 06-APR-2001; 2001US-0281732P.
 PP
 XX (ORIG-) ORIGENE TECHNOLOGIES INC.
 PA
 XX Sun Z, Jay G;
 PI
 XX WPI; 2003-058520/05.
 DR
 XX N-PSDB; ABX10341.
 XX Novel genes which are differentially regulated in prostate cancer, useful
 PT for diagnosing prostate cancer in prostate tissue sample and assessing
 PT therapeutic or preventive intervention in prostate cancer patients.
 XX
 XX Claim 1; Page 282-292; 416pp; English.
 PS
 XX The invention describes genes (I) which are differentially regulated in
 CC prostate cancer. (I) is useful for diagnosing a prostate cancer in a
 CC sample comprising prostate tissue, which involves determining the number
 CC of target genes which are differentially-regulated in the sample, where
 CC the number is indicative of the probability that the sample comprises
 CC prostate cancer. (I) is useful for assessing a therapeutic or preventive
 CC intervention in a subject having a prostate cancer, which involves
 CC determining the expression levels in a sample comprising prostate tissue
 CC of target genes which are differentially-regulated in prostate cancer.
 CC Preferably, the expression levels of at least 10 genes are determined.
 CC (I) is also useful for identifying agents that modulate a biological
 CC activity of a polypeptide differentially-regulated in prostate cancer
 CC cells, which involves contacting a polypeptide differentially-regulated
 CC in prostate cancer cells with a test agent under conditions effective for
 CC the test agent to modulate a biological activity of the polypeptide, and
 CC determining whether the test agent modulates the biological activity. (I)
 CC is useful as molecular markers, as drug targets, and for detecting,
 CC diagnosing, staging, grading, assessing, monitoring, prognosticating,
 CC preventing of treating, determining predisposition to diseases and
 CC conditions especially relating to prostate cancer. (I) and its expression
 CC products are used in the diagnostic test to assay for presence of cancer
 CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
 CC blood etc. (I) is useful for assessing cancer e.g., to determine the type
 CC of cancer, its stage of development, the nature of genetic defect, etc.
 CC The polypeptide encoded by (I) can be used as target for therapy or drug
 CC discovery. (I) can also be used for expressing the polypeptide and thus
 CC for searching specific binding partners of the polypeptide. (I) is useful
 CC in therapeutic applications to treat prostate cancer. The identification
 CC of specific genes, and groups of genes, expressed in pathways
 CC physiologically relevant to prostate cancer permits the definition of
 CC functional and disease pathways and the delineation of targets in these
 CC pathways which are useful in diagnostic, therapeutic, and clinical
 CC applications. This is the amino acid sequence of a protein differentially
 CC regulated in prostate cancer
 XX Sequence 3210 AA;

Query Match 3.7%; Score 187.5; DB 6; Length 3210;
 Best Local Similarity 20.1%; Pred. No. 0.071;
 Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;

QY 6 KTVALTITTVSVTHNQEVSLVKEPILKQTOASSISGADYAESGSKSLKINETSPV 65
 DB 2069 KTTALDQLSEKMKKEKTQEL-----ESHQSECLHCIQVAAEAVKTEKLTQT---L 2115
 QY 66 DDTVDLTFSDKRTTPEKI-----KDNLAGPREQBLKAVTENTESEKOTISGQLEQSKES 121
 DB 2116 SSDVSELLKDTHTLOEKLOSLEKDSQALSITKCELENIQIAQLNKEKELL-----VKES 2168
 QY 122 LSLNKTVPSTSNWEICD-----FITKGNLTVLGS-----KSGVEKLS---QTD 161
 DB 2169 ESLQARL-SESDEKLVNYSKALEAALVEKGSEFALRLSSQTOEVHQLRRGIEKLVRIEAD 2227
 QY 162 -----HLV--LPSQAADGTQLIQVASFAPTPDKKTAITABYTSRAGENGESQLDVG--- 211
 DB 2228 EKQQLHIAEKLKERERENDSL-----KDKYENLERELQMBENQELVILDAENSKA 2278
 QY 212 -----KEIINEGEVFNSYLLKKVTIPTGYKHI-----GQ-----DAF 243
 DB 2279 EVELTKQIEMARSLKVEF---LDLVTLRSEKENLTQIQEKGQQLSELDKLSSPKSL 2335
 QY 244 VDNKNIAEAVNLPESLETISDYAFALHALKQIDLPDNLKAIGELAPFDNQITGKLSLPRQL 303
 DB 2336 LEEKEQAEIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 2377
 QY 304 MRLAERAFKSNHIKTIEFRGNSLVIGEASFQDNLDLSQLMLPDGLEKIESEAFNGPDD 363
 DB 2378 MKATEQLDPPIEEHQLR-NSIEKL-RARLEADEKKQLCVLQQLKSEHHA-----D 2428
 QY 364 HYNRV-----VLMTSGKGNPSGLATENTVVPDKSLMQBSPEDIVTKMLEEDFTYQKN 417
 DB 2429 LLKORVENLERELEIARTNQEHAALEAENS-----KGEVETLAKIEGTQ--- 2474
 QY 418 SVTGFSNKGLO-----KVRNKNLEIPKOHNGVTTITEIGDNAFRNVDFQNKTLRYDL 470
 DB 2475 -----SLRGLELDVVTTIRSEKEDLTNLEKQEBRISLEIINSSFENI-LQEKEQKVOM 2528
 QY 471 BEVKLPSTIRKIGAFQSNILKSFASDDLEETK-EGAFNNRIETLELKDVLVTGDA 529
 DB 2529 KEKS--STAMEMLOTQLKELNERYAALHNDQEAQKQNLSSQVCELEKAQLLOGLD 2586
 QY 530 AFHINHIAIVLPSV---QEIGRSAPRONGANNLIFMGSKVKTGLGMAFLSNRLEHL 585
 DB 2587 EAKNY---IVLQSSVNGLIQEV-----EDGKKKL-----EKDEISLKKNIQI--- 2628
 QY 586 LSEQQLTEIPVQAFSDNAL-KEVLLPASLKTIRREAFKKNHLKQLEVASALSIAFNAL 644
 DB 2629 --DQQLVSKLSQVEGEHQLWKEQNL--ELRNLTFVELEQKIQVLQSKNAS-----L 2675
 QY 645 DDNDGDEQPDKNVVKVTHNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLR 704
 DB 2676 QD-----TLEVLQSSYKNLENELELTQMDKMSF-----VEKNKM-----TAK 2713
 QY 705 QT-TQTFQDMTTAGKALLSKNSLRQCEKQFLQEAQFFLGRVLDLDAIAKAELVTK 763
 DB 2714 ETELQREHMEH--AQKTAEIQEEL-SEKRNLAGELQLLEEI-----K 2754
 QY 764 ATKNGQLLERSINKAVLAYNNNAIKKA-----NVKRLKEKLDLLTGLVEGKPLAQATMVQ 819
 DB 2755 SSKD-QLKE-----LTLENSELKSLDCHMKQKQVEKGVREIEAYQLRHEAEKKH 2806
 QY 820 GYLLKTPPLPPEYIYGLNVYFDK-----SGKLIYALDMSDTIG 858
 DB 2807 QALLLDTN---KQYEVEIQTREKLTSKBECCLSQKLEIDLLKSKKEELNLSKATQIL 2863
 QY 859 EGQKDA-----YGNPILNVDEDNQVHALAVATLADYEGLDIKTILNLSQL----- 906
 DB 2864 EELKTKMDNLKYNVQLKKNERAGQKKLLIKSKQLE--BEKEILQKELSQAQAEK 2921
 QY 907 -----TSIRQVPTAAYHRAGIFQAIQNAABAEQOLLPKPG-----THSEKSSSESAN 954
 DB 2922 QKTGVMDTKVDELTTEIKE---LKETLEEKTKEADEYLDKYCSLLISHEKLEKAKEMLE 2978

QY	955 SKDRGLQSNPKTNRGHSAIL 975	QY	6 KTVALTTLTVSVVTHNQVFLSKVPEILKQTOASSISSGADYAEBSGSKLKNINETSQPV 65
Db	2979 TQVAHLCSQSQSKDSRGSPILL 2999	Db	2069 KTTALDQLSEKMKETQEL-----ESHQSECLHICIQVAEAEVKEKTELLQT---L 2115
RESULT 97		QY	66 DDTVTDLPSDKRTTPEKI-----KONLAKGPREQELKAVTENETSEKQITSGSQLEQSKES 121
ADJ66483		Db	2116 SSDVSELKDKTHLOEKLOSKLEKQSQALSILTKCELENIQIAOLNKEKELL-----VKES 2168
ID	ADJ66483 standard; protein; 3210 AA.	QY	122 LSLNKTVPSTSNWEICD-----FITKGNLTVLGLS-----KSGVEKLS---QTD 161
AC	ADJ66483;	Db	2169 ESLQARL--SESDYEKLVNSKALEAALVEKGEFALRLSTQEEVHQLRGIEKLRVLEAD 2227
DT	06-MAY-2004 (first entry)	QY	162 -----HLV--LPSQAADGTQLIQVASFAPTPDKTAIAEYTSRAGENGEISQLDVG--- 211
DE	CenP-F kinetochore protein for anti-cancer protein complex.	Db	2228 EKKQLHIAEKLKERENDSL-----KDKVENLERELQMEENQELVILDAENSKA 2278
KW	neuroprotective; cytostatic; gene therapy; protein complex;	QY	212 -----KEIINEGEVFNYSLLKKYTIPTGYKHI-----GQ-----DAF 243
KM	cellular network; cancer; neurodegenerative disease; drug target.	Db	2279 EVETLKTQIBEMARSLKVF-----LDLVTLRSEKENTLTKQIQKQGLSELDKLLSSPKSL 2335
XX	Homo sapiens.	QY	244 VDNKNIAEVNLPESLETISDYAFALHALKQIDLPNLKAIGELAPFNOITKGLSLPQL 303
OS	WO2004009622-A2.	Db	2336 LEEKQAEIQIEESKTAVEM-----LNQQLKEL-----NEAVALCGQOEI 2377
XX	29-JAN-2004.	QY	304 MFLAERAFKSNHIKTIEFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEAFTPGNPDD 363
XX	18-JUL-2003; 2003WO-EP007835.	Db	2378 MKATEQSLDPPIEERHQLR-NSIEKL-RARLEADEKKQLCVLQQLKESEHHA-----D 2428
XX	19-JUL-2002; 2002EP-00016109.	QY	364 HYNRRV-----VLMTKSGKNPGLATENTYVNPDKSLWQESPIDYTKLWLEEDFTYQKN 417
XX	19-JUL-2002; 2002EP-00016111.	Db	2429 LLKGRVENLERELEIARTNOEHALEAENS-----KGEVETLKAKIEGTMQ--- 2474
XX	19-JUL-2002; 2002EP-00016123.	QY	418 SVTGFSGKGLQ-----KVKRNKNLEIPKOHNGVITTEIGDNAPRVDFQNTLRKVDL 470
XX	22-JUL-2002; 2002EP-00016128.	Db	2475 -----SLRGLDLVVITRSEKEDLTNLOKEQERISELEIINSFENI--LOEKEQEKVQM 2528
XX	2002EP-00016427.	QY	471 BEVKLPSTIRKIGAFQFOSNNLKSFEASDDLLEIK-EGAFMNRNRIETLEKDKLVITGDA 529
PA	(CELL-) CELLZONE AG.	Db	2529 KEKS--STAMEMLOTQELNERNVAALHNDQEAQKAEQNLSQVECLELEKQLLQGLD 2586
XX	Merino A, Bouwmeester T, Bauer A, Drewes G, Marzioch M, Kruse U;	QY	530 AFHINHIYAVLPSV-----QEIGRSAPRQNGANNLIFPMGSKVTGLGEMAFNLRLHLD 585
PI	Superti-Furga G, Eberhard D, Ruffner H, Hobson S, Helftenbein G;	Db	2587 EAKNNY-----IVLQSSVNGLIQEV-----EDGKQL-----EKKDEIISRLKNQIQ--- 2628
PI	Cruciat C;	QY	586 LSEQKQLTEIPVOAFSDNAL--KEVLLPASLKTIREEAFKKNHLKQLEVASALSHIAPNAL 644
DR	WPI; 2004-123372/12.	Db	2629 --DQQLVSKLSQVEGEHQLWKEQL--ELRNLTVELEQKIQLVQSKNAS-----L 2675
XX	New protein complexes of cellular networks underlying the development of	QY	645 DNDGDGEQFNDKVVVKKVTHNSYALADGEHFIVDPDKLSSTIVDLKILKILGIDYSTLR 704
PT	cancer and other diseases, useful for diagnosing and/or treating	Db	2676 QD-----TLEVLQSSYKLENELELTWKMDKMSF---VEKYNKM-----TAK 2713
PT	neurodegenerative diseases or cancer, and in drug screening.	QY	705 QT-TQTQPRDMMTAGKALLSKNSLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKK 763
XX	Disclosure; SEQ ID NO 13; 809pp; English.	Db	2714 ETELQREHMHM--AQKTAELQDEL--SGEKNRLAGELQLLEEI-----K 2754
XX	The invention relates to a protein complex of cellular networks	QY	764 ATKNGQLERSINKAVLAYNNSAIKKA-----NVKRLKELDLTLGLVEGKPLAQATMVQ 819
CC	underlying the development of cancer and other diseases. The complex (I)	Db	2755 SSKD-QLKE-----LTLENSELKSLDCHMDKQVEKGVREIETAEYQRLRHEAKKH 2806
CC	comprises at least one first and second proteins selected from any of the	QY	820 GYLLKTLPLPEYVIGLVNLYVFDK-----SKGLYALDMSDTIG 858
CC	proteins listed in the specification, or their functionally active	Db	2807 QALLLDTN---KQYEVEIQTREKLTSEKCLSSQKLEIDLKSSKELANNSLKATTOIL 2863
CC	derivatives, fragments, homologues or variants, the variants being	QY	859 EQQKDA-----YGNPILNVDDNEGHALAVATLADYEGDLDTIKTLNLSKLSQJ----- 906
CC	encoded by a nucleic acid that hybridizes to the nucleic acid encoding	Db	2864 BELKTKMDNLKYNVNLKKNERAAQGMKLLIKSCQLE--EKEIILQKELSQLQAAQEK 2921
CC	the protein under low stringency conditions. A complex (II) comprises at	QY	907 -----TSIROVPTAYHRAGIFQAIQNAAAAEQLLPKG-----THSEKSSSESAN 954
CC	least two of the second proteins, where the low stringency conditions	Db	2922 QKTGTGMDTKVDELTEITKE---LKETLEETKEADBYLDKYCSLLISHSEKLEKEMLE 2978
CC	comprise hybridization in a buffer comprising 35% formamide, 5 x SSC, 50		
CC	mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% pVP, 0.02% BSA, 100 microg/ml		
CC	denatured salmon sperm DNA, and 10% (wt/vol) dextran sulfate for 18-20		
CC	hours at 40 deg C, washing in a buffer consisting of 2 x SSC, 25 mM Tris-		
CC	HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS for 1.5 hours at 55 deg C, and		
CC	washing in a buffer consisting of 2 x SSC, 25 mM Tris-HCl (pH 7.4), 5 mM		
CC	EDTA, and 0.1% SDS for 1.5 hours at 60 deg C. The composition and methods		
CC	are useful in diagnosing or treating diseases and disorders, preferably		
CC	neurodegenerative diseases. These may also be used as a drug target or in		
CC	manufacturing a medicament for the treatment or prevention of the above-		
CC	mentioned diseases or disorders. The composition may also be used for		
CC	treating cancer. This sequence represents one of the proteins of the		
CC	complex of the invention.		
XX	Sequence 3210 AA;		
QY	Query Match 3.7%; Score 187.5; DB 8; Length 3210;		
XX	Best Local Similarity 20.1%; Pred. No. 0.071;		
XX	Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;		

```

QY 955 SKDRGLQSNPKTNRGRHSAIL 975
Db 2979 TOVAHLCQQSQKQDSRGSPLL 2999
RESULT 98
ID ABB63571
XX ABB63571 standard; protein; 1348 AA.
AC ABB63571;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 17505.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers BW;
XX WPI; 2001-658660/75.
DR N-PSDB; ABL07674.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 17505; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1348 AA;
XX
XX Query Match 3.7%; Score 187; DB 4; Length 1348;
XX Best Local Similarity 18.6%; Pred. No. 0.023;
XX Matches 219; Conservative 159; Mismatches 344; Indels 454; Gaps 55;
XX
XX 92 REQELKAVNTESKQITSGSQSKESLSLN-KTVPSTSNWEICDFTKGNLTGLS 150
XX 11 RLQRELAKQSPROKQKTAKLPSSQATKSLKCNQAAPKTET-----ENTF-QGL 60
XX
XX 151 KSGVKLSQTDHLV--LPSQAADGTLIQVASFAPTPDKTAIABYTSRAGEN----- 201
XX 61 KLTIELDLVNLIRRIPEKAFDGLK-----DSLNELRNLNLLGDNLPFST 109
XX
XX 202 -----GISQLDVQKEI--INEGEVFNYSLLKVTIPTGYKHIGQDAFVFNKNIARV- 252
XX 110 AELHVLKLNRLDLSGNKIKLIEEG-----LLK-----GCMDL-KEFYIDRNSLTSPV 156
XX
XX 253 ----NLPSLETISDYAFALHAKQIDLPDNLKAIGELAFPDNQITGKLSLPQLMRLAE 308
XX

```

RESULT 99
 ABB77984
 ID ABB77984 standard; protein; 1312 AA.
 XX
 AC ABB77984;
 XX

DT 22-OCT-2002 (first entry)
 XX Amino acid sequence of a yeast RAD50 protein.
 DE
 KW Nucleic acid integration; homologous recombination; telomeric region;
 KW RAD50.
 XX
 XX Saccharomycetes cerevisiae.
 XX
 XX EP1217074-A1.
 XX
 XX 26-JUN-2002.
 XX
 XX 22-DEC-2000; 2000EP-00204693.
 XX
 XX 22-DEC-2000; 2000EP-00204693.
 XX
 XX (UYLE-) RIJKSUNIV LEIDEN.
 XX (BINA-) STICHTING BINAIR VECTOR SYSTEEM.
 XX
 XX Hooymaas PJJ, Van Attikum H, Bundock P;
 XX WPI; 2002-550409/59.
 XX
 XX Directing integration of nucleic acid of interest to a sub-telomeric
 PT region in an eukaryote with preference for non-homologous recombination,
 PT by steering an integration pathway towards homologous recombination.
 XX
 XX Disclosure; Fig 5; 63pp; English.
 XX
 XX The specification describes a method for directing integration of a
 CC nucleic acid of interest to a pre-determined site, where the nucleic acid
 CC has homology at or around the pre-determined site, in a eukaryote with a
 CC preference for non-homologous recombination. The method comprises
 CC steering an integration pathway towards homologous recombination. The
 CC method is useful for directing integration of a nucleic acid of interest
 CC to a subtelomeric and/or telomeric region in an eukaryote with a
 CC preference for non-homologous recombination. The nucleic acid of interest
 CC comprises an inactive gene to replace an active gene, or vice versa, is a
 CC portion of a gene delivery vehicle, confers a desired property to the
 CC eukaryotic cell, or encodes a therapeutic proteinaceous substance or a
 CC substance conferring resistance for an antibiotic substance to a cell.
 CC The method is useful for improving gene targeting efficiency. The method
 CC is useful in the replacement of an active gene by an inactive gene, for
 CC e.g. for the inactivation of genes controlling undesired side branches of
 CC metabolic pathways, to increase the quality of bulk products such as
 CC starch, or to increase the production of specific secondary metabolites
 CC or to inhibit formation of unwanted metabolites, and also to inactivate
 CC genes controlling senescence in fruits and flowers or that determine
 CC flower pigments. The method is also useful for replacing an inactive gene
 CC by an active gene. For e.g. the replacement of a defective p53 by an
 CC intact p53. Many tumours acquire a mutation in p53 during their
 CC development which renders it inactive and often correlates with a poor
 CC response to cancer therapy. By replacing the defect p53 by an intact p53,
 CC e.g. through gene therapy, conventional anti cancer therapy have better
 CC changes of succeeding. The method is also useful for therapeutic
 CC proteinaceous substance integration. A tumoricidal gene can be delivered
 CC to a pre-determined site present only in e.g. proliferating cells, or
 CC present only in tumour cells, e.g. to the site where a tumour antigen is
 CC expressed form. ABB77984-86 represent RAD50 homologues. RAD50 is involved
 CC in non-homologous recombination
 XX
 SQ Sequence 1312 AA;
 Query Match 3.7%; Score 186.5; DB 5; Length 1312;
 Best Local Similarity 19.0%; Pred. No. 0.024;
 Matches 200; Conservative 174; Mismatches 362; Indels 315; Gaps 49;
 QY 21 NQEVSLVKEPIKQTOASSISGADYAESGSKLKNINETSQPVDDTVDLPSKRTTP 80
 DB 361 NRRHLSLKEAFQHKPQGLSINSDMAQVNHMSQFKAFISQDLTDIDQFAKDIQLKE 420
 QY 81 EKIKD-----NLAKGPREQ-----ELKAVTENTESEKQITSGSQSKESLSLN 125

Db 421 TNLSDLIKSITVDSQNLVYKDKRSKLHDSLEAEKLSFKSLSTQDSLHLENLTKY 480
 QY 126 KTVPTSTNWEICDFITKGNLTVG-----LSKSGVEKLSQTDHLVLSQAD----- 171
 Db 481 K--EKLQSWESNIIPKLNQKIEKNEMILENOIERFQ--DRIMKTNQADLVAKUGL 536
 QY 172 -----GTQLIQVASPAFTPDKKTAIAEYTSRAGEGEISQ-----LDVDGKEI--INEGE 219
 Db 537 IKKSINTKLDEL-----QKITEKLQNDLSRIQVPLTQEFQORADLEMDFKLFIN--- 586
 QY 220 VNSYLLKKVITPTGYKH-----IQDAFVNDKNTAEV-----NLPSL 258
 Db 587 -----MQKNIAINNNKQMHEDRRYTNALYNLNTIEKDLQDNQKSKERVTQLSENLPDC 641
 QY 259 ETISDY-----APAHLLAKOIDLPDNLKAJGELAFFNQITGKLSLPQLMR 305
 Db 642 -TIDBYNDVLESTELSYKTALENLKHQHTTLEFNKAL--EIAERDS-----CCY 688
 QY 306 LAERAFKSNHKTIFRGNLSKVIGEASFQ-----DNLSQLMLPDGLEKIESEAFGNPG 361
 Db 689 LCSRKFENESFKSLQ--ELKTKTDANFEKTLKDTVQNEKEYLHSLALLEKHIIITLSI 746
 QY 362 DDHVN-----RVVLWTKSGKNP--SGLATENTYVNDPKSLWQES--FEIDYTKMLBEDF 412
 Db 747 NEKIDNSQKCLEKAEETKTSKLDLEVDSTKLKDEKELAESEIRPLIERFTYLEKEL 806
 QY 413 TVQKN-----SVTGFSNKGLO-----KVK 431
 Db 807 KDLNSSWTISELSYNTSDEGIQTVDELDRDQQRQMDLSRELKRTISDLQMEKDKVR 866
 QY 432 RNKMLEIPKQNGVITTEIGONAF--RNVDFOFKTLRYDYLEEVK--LPSTIRKIGAFAPQ 488
 Db 867 ENSRMNLIKELTVSIESSLTKQKIDDSIRSKR-----ENINDIRVKELEARIIS 922
 QY 489 SNLKSFPASDDLEIKGAPMNNRIETLEKDKLVTTGDAFHNHI---YAVILPESV 545
 Db 923 LKNKKD-BAQSVLDKVK-----NER--DIQVRNKQKTVAD-----INRLIDRFQTYNEV 970
 QY 546 QEIGRSAPQNCANNLI FMGSKVKTGLGMAFLSNRLEHLDLSEOKQLTEIPVQAFSDNAL 605
 Db 971 D-----PEAKGFDEL-----QTTIKLEL--NKAQMLELKEQDLKSNV-----NEE 1011
 QY 606 KEVLLPASLTKITREAPFKNNHKLQLEVASALSHIAFNALDNDGDGEQFDNKVVVKTHNS 665
 Db 1012 KRKLADSN-----NEEKNLKNLELLELKSQLOHI-----ESE 1044
 QY 666 YALADGEHFI VDPDKLSSTIVDLKILKILEGLDYSTLRQTQTQFDMWT--AGKALLS 723
 Db 1045 ISRLDVQNAEARDKYQ-----EESRL-----RTRFEKLSSENAGKL--- 1082
 QY 724 KSNLRQGEKQKFLQEAQFGLGRVDLDKAIKAEKALVTKK--ATKNGQLLERSINKAVLAY 782
 Db 1083 -GEMKQLQNLQIDSLTHQLRTDYKIEKNYHKEWELQTRSFVTDIDIVYSKALDSAIMKY 1141
 QY 783 NNSATKKANVRKLEKEL-----DLLTGLVEGKGFLAQATMVQGVYLIK 825
 Db 1142 H--GLKMQDINRIIDELWKRTYSGTIDITKIRSDVSSTVKGSYNYRVVMYK----- 1193
 QY 826 TPLPLPEYIYGLNVYFDKSGK-----LIYALDMSDTITGEG--QKDAYGNPILNVD 874
 Db 1194 -----QDVELDMRGCSAQKVLASIIIRLALSETFGANGCVIALDEPTTNLDE 1242
 QY 875 DNEGHALAVATLADYEGLDIKTILNSKLSQ 905
 Db 1243 EN-----IESLAK-----SLHNIINMRHQ 1262
 RESULT 100
 ABR53887
 ID ABR53887 standard; protein; 1312 AA.
 XX
 AC ABR53887;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:37:35 ; Search time 48 Seconds
(without alignments)
2020.552 Million cell updates/sec

Title: US-10-078-531-2
Perfect score: 5080
Sequence: 1 MKKHLKTVALLTITVSVVTH.....LGYTSVALLSLITAIKKKKY 1008

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 110 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	331.5	6.5	1081	2	T31094
2	221	4.4	690	2	P98114
3	219	4.3	1208	2	T39068
4	217.5	4.3	1072	2	A86827
5	214.5	4.2	3225	2	I52300
6	214.5	4.2	3259	1	A56539
7	213.5	4.2	621	2	A95250
8	212.5	4.2	3216	2	C90538
9	209.5	4.1	1227	2	C97033
10	209.5	4.1	1252	2	B42771
11	208.5	4.1	2829	2	A42771
12	206.5	4.1	2269	2	T28677
13	204.5	4.0	2481	2	D90011
14	204.5	4.0	4319	2	T31105
15	203.5	4.0	2346	2	T13829
16	203	4.0	6713	2	B89921
17	199.5	3.9	2285	2	T12796
18	199	3.9	2186	2	H89960
19	197.5	3.9	1269	2	F84730
20	197.5	3.9	1790	2	S67593
21	197	3.9	1119	2	H70126
22	197	3.9	1127	2	T28317
23	197	3.9	2441	2	D71623
24	196.5	3.9	1875	2	S38173
25	196.5	3.9	2139	2	T18296
26	195.5	3.8	1778	2	AF1116
27	194.5	3.8	1108	2	T18353
28	194	3.8	949	2	E71940
29	191	3.8	1365	2	T30822

30	190.5	3.8	1727	2	T50073
31	190.5	3.8	3488	2	T34418
32	189.5	3.7	925	2	G88175
33	189.5	3.7	1957	2	T38077
34	188.5	3.7	2013	2	AD1129
35	186.5	3.7	1092	2	T18354
36	186.5	3.7	1312	1	BWBYDL
37	185.5	3.7	890	2	S22452
38	185.5	3.7	1186	2	G69708
39	185	3.6	1642	2	T08880
40	183	3.6	1091	2	T34107
41	182.5	3.6	944	2	H64650
42	182.5	3.6	1313	2	F96673
43	182	3.6	1173	2	T43527
44	182	3.6	1622	2	AB1717
45	182	3.6	1946	2	AE1449
46	182	3.6	4151	2	T13734
47	181.5	3.6	1751	2	A45604
48	181.5	3.6	2026	1	OVBV
49	180.5	3.6	1959	2	AG1085
50	180.5	3.6	1992	1	S02771
51	180	3.5	1302	1	JC6009
52	178.5	3.5	2819	2	A90551
53	178	3.5	891	2	G41662
54	177.5	3.5	1427	2	S22695
55	177.5	3.5	1817	2	H71611
56	177	3.5	2401	2	T28676
57	177	3.5	2712	2	T05113
58	176.5	3.5	1237	2	AC1583
59	176	3.5	1170	2	A72287
60	176	3.5	1431	2	A45866
61	175.5	3.5	1191	2	B97116
62	175.5	3.5	1619	2	S67083
63	175	3.4	1099	2	T18357
64	175	3.4	1151	2	AG1717
65	175	3.4	1392	2	A43336
66	174.5	3.4	976	2	T01553
67	174.5	3.4	1690	2	T13030
68	174	3.4	1558	2	B71603
69	174	3.4	1744	2	F86161
70	174	3.4	1839	1	OVBK
71	174	3.4	2052	2	C97038
72	173.5	3.4	2748	2	S57976
73	173.5	3.4	2954	2	T14156
74	173	3.4	1079	2	T30996
75	173	3.4	1332	2	S41552
76	173	3.4	1819	2	D97033
77	172	3.4	1066	2	B95037
78	172	3.4	1066	2	T15864
79	172	3.4	1189	2	A54817
80	172	3.4	1385	2	D89824
81	172	3.4	4688	2	F82885
82	171.5	3.4	1822	2	S34441
83	171.5	3.4	1939	2	D97316
84	171.5	3.4	2253	2	T30336
85	171	3.4	1368	2	T18371
86	171	3.4	1744	2	JH0720
87	171	3.4	1871	2	D86796
88	171	3.4	1963	2	B98002
89	171	3.4	2116	2	A26545
90	170.5	3.4	918	2	G88545
91	170.5	3.4	1620	2	S61535
92	170	3.3	943	2	S31132
93	170	3.3	992	2	S54396
94	170	3.3	1864	2	F86378
95	170	3.3	1928	2	S46773
96	169.5	3.3	1169	2	A64505
97	169.5	3.3	1881	2	H95076
98	169	3.3	829	2	T01362
99	169	3.3	1051	2	T13174
100	169	3.3	2385	2	A32491
101	169	3.3	2484	2	T26216
102	169	3.3	2607	2	T26215

myosin-like coiled
hypothetical prote
protein T24H7.2 (i
hypothetical coile
probable peptidogl
adhesin - Mycoplas
RAD50 protein - ye
surface exclusion
chromosome segrega
NMDA receptor-bind
hypothetical prote
translation initia
hypothetical prote
sp8 protein - firs
probable cell surf
hypothetical prote
groovin gene prote
major blood-stage
adenylate cyclase
hypothetical prote
myosin heavy chain
surface-located me
conserved hypothet
130K surface exclu
resistin - human
probable secreted
rhopty protein -
hypothetical prote
internalin protein
hypothetical prote
dextranucrase (EC
chromosome segrega
hypothetical prote
mhl protein - Myc
probable peptidogl
microtubule-vesicl
hypothetical prote
microtubule bindin
RESA-H3 antigen pF
F1003.10 protein -
adenylate cyclase
phage-related prot
nuclear migration
kinesin-related pr
hypothetical prote
probable transcrip
uncharacterized pr
hyaluronidase [imp
hypothetical prote
ATPase ScII, chrom
hypothetical prote
EF protein - Strept
probable S-layer p
nuclear/mitotic ap
probable glutamate
tanabin - African
probable heat shoc
IGA-specific metal
myosin heavy chain
protein F59B2.12 (i
nucleotide-binding
protein L precursor
protein F21J9.12 (i
myosin heavy chain
P115 homolog - Met
zinc metalloprotei
probable myosin he
gp150 protein - fr
myosin heavy chain
hypothetical prote
hypothetical prote

103 168 5 3.3 1295 2 T24587
104 168 5 3.3 1639 2 S05603
105 168 5 3.3 2491 2 A57036
106 168 3.3 909 2 C97325
107 168 3.3 1090 2 H86806
108 168 3.3 1466 2 A36426
109 168 3.3 2144 2 A97942
110 168 3.3 2411 2 B32491

hypothetical prote
major mezozoite su
talin - slime mold
hypothetical prote
hypothetical prote
SPA2 protein - yea
metalloproteinase
myosin heavy chain

ALIGNMENTS

RESULT 1
T31094
surface antigen BepA - Bacteroides forsythus
C:Species: Bacteroides forsythus
C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004
C:Accession: T31094
R/Sharma, A.; Sojar, H.T.; Glurich, I.; Honma, K.; Kuramitsu, H.K.; Genco, R.J.
Infect. Immun. 66, 5703-5710, 1998
A>Title: Cloning, expression, and sequencing of a cell surface antigen containing a leuc
A:Reference number: Z20977; MUID:99043895; PMID:9826345
A:Accession: T31094
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1081 <SHA>
A:Cross-references: UNIPROT:O68831; EMBL:AF054892; NID:g3005673; PIDN:AA08
C:Genetics:
A:Gene: bcpA

Query Match 6.5%; Score 331.5; DB 2; Length 1081;
Best Local Similarity 25.3%; Pred. No. 1.1e-07;
Matches 147; Conservative 83; Mismatches 206; Indels 144; Gaps 23;

Qy 110 TSGSLEQSKESLSN-----KTPSTSNWEICDFITKNTLVGLSKSGVEKLSQTDH 162
Db 19 TLGATAQNSGTTGPLNWSYDSGKTALITGCMPDF-----NNASEIPWHLSQSKIQT-- 72
Qy 163 LVLPSQAADGTQLIQVAFPT-----PDKKTAIAEYTSRAGENGEISQLDVGKE 213
Db 73 -----VTIGDGVTSVGNFASDCALTSVTLPSNLTAIGDHAFK-GCSGLTSIITPNSVT 125
Qy 214 IINEGEVNSYLLKKVTPTGKYKHQDAFVDNKNIAEVNLPESLETISDYAPAHLA-LK 272
Db 126 TIGWAFKGCGLSKSITLPSNLTAIGSALSGCTGLTSITIPNSVTITIGWAFKCSGLT 185
Qy 273 QIDLPDNLKAIGELAFFDNOITGKLSLPRQLMRLAERAFKS-NHIKTEFRGNSLKVIGE 331
Db 186 SITFNSLTAIGESAFYCGGALTSTLPDALTTIGESAFKCGSLKSIPTF-PNSLTTIGE 244
Qy 332 ASFQD-NDLSQLMPLDGLIESEAFNPGDGHYNNRVLWTKSGKNPSGLATENTYVN 390
Db 245 SAFVDCGALTSTLPDALTTIGRSFYG-----CSGL----- 276
Qy 391 PDKSLMQESPIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPKQHNGVTITEI 450
Db 277 -----KSIITFPN-----SLTTI 288
Qy 451 GDNAFRNVDFQNKTLRKYLDEVLKLPSTIRKIGAFAPQS-NNLKSFEASDDLBEIKGAF 509
Db 289 GESAFYNGC-----SLTSITIPNSVTTIGRSAFYCGSLKSIITLPDGLTTIEERAP 339
Qy 510 MNNRIET-LEIKDLKLVITGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGANNLIPMGSK 567
Db 340 YNCGLVTSITIPNSVATIGESAFYCGSLKSIITLPDGLTTIEWGAFYNGCALTSTIPNS 399
Qy 568 VKTIGEMAFLS-NRLEHLDLSEQKLTPIPVQ--AFSDNALKEVLL--PASLKTIREEAF 622
Db 400 VSTIGESAFYCGGALKQDVTVAMD---TPIDIQRDVFRELTLSGIRLHVPAKKTVE--- 453
Qy 623 KKNHLKQLEVASALSHIAFNALDDND-GDEQFDNKVVKVT 661

103 168 5 3.3 1295 2 T24587
104 168 5 3.3 1639 2 S05603
105 168 5 3.3 2491 2 A57036
106 168 3.3 909 2 C97325
107 168 3.3 1090 2 H86806
108 168 3.3 1466 2 A36426
109 168 3.3 2144 2 A97942
110 168 3.3 2411 2 B32491

hypothetical prote
major mezozoite su
talin - slime mold
hypothetical prote
hypothetical prote
SPA2 protein - yea
metalloproteinase
myosin heavy chain

ALIGNMENTS

RESULT 1
T31094
surface antigen BepA - Bacteroides forsythus
C:Species: Bacteroides forsythus
C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004
C:Accession: T31094
R/Sharma, A.; Sojar, H.T.; Glurich, I.; Honma, K.; Kuramitsu, H.K.; Genco, R.J.
Infect. Immun. 66, 5703-5710, 1998
A>Title: Cloning, expression, and sequencing of a cell surface antigen containing a leuc
A:Reference number: Z20977; MUID:99043895; PMID:9826345
A:Accession: T31094
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1081 <SHA>
A:Cross-references: UNIPROT:O68831; EMBL:AF054892; NID:g3005673; PIDN:AA08
C:Genetics:
A:Gene: bcpA

Query Match 6.5%; Score 331.5; DB 2; Length 1081;
Best Local Similarity 25.3%; Pred. No. 1.1e-07;
Matches 147; Conservative 83; Mismatches 206; Indels 144; Gaps 23;

Qy 110 TSGSLEQSKESLSN-----KTPSTSNWEICDFITKNTLVGLSKSGVEKLSQTDH 162
Db 19 TLGATAQNSGTTGPLNWSYDSGKTALITGCMPDF-----NNASEIPWHLSQSKIQT-- 72
Qy 163 LVLPSQAADGTQLIQVAFPT-----PDKKTAIAEYTSRAGENGEISQLDVGKE 213
Db 73 -----VTIGDGVTSVGNFASDCALTSVTLPSNLTAIGDHAFK-GCSGLTSIITPNSVT 125
Qy 214 IINEGEVNSYLLKKVTPTGKYKHQDAFVDNKNIAEVNLPESLETISDYAPAHLA-LK 272
Db 126 TIGWAFKGCGLSKSITLPSNLTAIGSALSGCTGLTSITIPNSVTITIGWAFKCSGLT 185
Qy 273 QIDLPDNLKAIGELAFFDNOITGKLSLPRQLMRLAERAFKS-NHIKTEFRGNSLKVIGE 331
Db 186 SITFNSLTAIGESAFYCGGALTSTLPDALTTIGESAFKCGSLKSIPTF-PNSLTTIGE 244
Qy 332 ASFQD-NDLSQLMPLDGLIESEAFNPGDGHYNNRVLWTKSGKNPSGLATENTYVN 390
Db 245 SAFVDCGALTSTLPDALTTIGRSFYG-----CSGL----- 276
Qy 391 PDKSLMQESPIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPKQHNGVTITEI 450
Db 277 -----KSIITFPN-----SLTTI 288
Qy 451 GDNAFRNVDFQNKTLRKYLDEVLKLPSTIRKIGAFAPQS-NNLKSFEASDDLBEIKGAF 509
Db 289 GESAFYNGC-----SLTSITIPNSVTTIGRSAFYCGSLKSIITLPDGLTTIEERAP 339
Qy 510 MNNRIET-LEIKDLKLVITGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGANNLIPMGSK 567
Db 340 YNCGLVTSITIPNSVATIGESAFYCGSLKSIITLPDGLTTIEWGAFYNGCALTSTIPNS 399
Qy 568 VKTIGEMAFLS-NRLEHLDLSEQKLTPIPVQ--AFSDNALKEVLL--PASLKTIREEAF 622
Db 400 VSTIGESAFYCGGALKQDVTVAMD---TPIDIQRDVFRELTLSGIRLHVPAKKTVE--- 453
Qy 623 KKNHLKQLEVASALSHIAFNALDDND-GDEQFDNKVVKVT 661

Db 454 AKDVWKE-----FNIVEDDDFGGLQWNYDAATKT 482

RESULT 2
P98114
choline-binding protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: P98114
R:Hoskins, J.A.; Arnold Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Mateuaghima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: P98114
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-690 <KUR>
A:Cross-references: UNIPROT:Q8DN38; GB:AE007317; PIDN:AA00747.1; PID:g15459643; GSPDB:G
C:Genetics:
A:Gene: pcpA

Query Match 4.4%; Score 221; DB 2; Length 690;
Best Local Similarity 23.2%; Pred. No. 0.007;
Matches 134; Conservative 71; Mismatches 177; Indels 196; Gaps 29;

Qy 6 KTVALTITTVSVW---THNQEVPS--LVKEPILKQTOAGSSI----- 42
Db 3 KTTLSLTTAAVILAAVYVPEPILAAVYVPEPILADTPSSEVIKVKVSGIIQNNIKYK 62
Qy 43 -----SGADYAESGSKLKNETSGPVDVDTVDLFS----- 74
Db 63 VLTVEGNIGTVQGVGTPVFEAGQDKPPTIPTKITVGDKVTVTEVASQAFSYVDE 122
Qy 75 -----DKRTTPEKIKDNLAKGPREQELKAVTENTSEKQITSGSQLEOSKE----- 120
Db 123 TGRIVVYPSIITPSSIKKIQKGFHSGKAKTII-----FDKGSQLEKIEDRAFDFS 174
Qy 121 -----SLSLNK-TVPSTSNWEICDFITKNTLVGLSKSGVEKLS 158
Db 175 ELEIEIPLASLEYIGTSAPFSQKLKLTPTSSSSKLEL-----ISHEAPANLS 222
Qy 159 QTDHLVLPQAAD-GTOLIQA-----SPA-----FTPDKKTAIAEYTSRA 198
Db 223 NLEKLTLPKSVKTLGSLNLFRLTSLKHVDVEEGNESPASVDGLVFSKD-KTQLIYPSQ- 280
Qy 199 GENGESIQLDVDGKEIINEGEVFNPSYLLKKVTIPTGKYKHQDAFVDNKNIAEVNLPESL 258
Db 281 -KNDES YKTPKTKELASYSFNKNSY-LKKLELNEGLEKIGTFAPADAIKLEELSLPNSL 338
Qy 259 ETISDYAP-AHLAKQIDLPDLNKAIGE-----LAPFDNOITGKLSLPRQLMRLAE 308
Db 339 ETIERLAFYGNLELKEILLPDNKNFKGVWNGLPKFLTLSGNNIN---SLPSFPLSVL 395
Qy 309 RAFPNSHIK-----TIEFRGNS--LKVIGESAFQ-DNDLSQMLPDGL 348
Db 396 DSLKEIHIKNKSTFSSVKKDTFAIPETVKFVTVSEHIKDLKSNLSTNDI-----IVEKV 451
Qy 349 EKISEAFTGNPGDDHYNNRVVLWTK-----SGKNPSGLATENTYVNPDKSLW-- 396
Db 452 DNIKQETDVAKP-KKNSNQGVGVWVKDKGLWYLYNESGSMATGWV-----KDKGLWY 503
Qy 397 -QESPEIDYTKWLEED--FTYQKNS---VTGF-SNKG 427
Db 504 LNESGSM-ATGWVKDKGLWYLYNESGSMATGWVKDKGL 540

RESULT 3
T39068
coiled coil protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe


```
QY 454 AFRNVDFONKTLRKYDLEEVKLPSTIRKIGAFAPFOSNNLKSFEASD-----DLEE 503
Db 332 TLNLSISLSDSSISQSTENQ--SGASSTAIRISYDSSENSQNSNSSEKSDNQ 389
QY 504 IKEGAFMNNRIETLELKDVLKLTIGDAAPIHNIHVAIVLP--ESVQETGRSAFRQNGANN 560
Db 390 SSLGSSMSNBSHNSNSNETNNSSEITNLPSPNPTESNVSDQTSSEASTNSNS 449
QY 561 LIFMGSKVKTIGEMAFNLRLHLDLSEQKQLTEIPVOAFDNALKEVLLPASLKTRIBE 620
Db 450 ISLSPSNISSTSDSESATNSDFSNVAEVANNSLASVNNSSSVLSSTADNL-GINQS 508
QY 621 AFKNHKLQLEVASALSHIAFNALDDNDGDQFNDKVVVKTTHNSYALADGEHFIYDPDK 680
Db 509 GSDNLTKDSSEISTSGAFLSNQTSS-----ASTNSNS-----SISLSPN 550
QY 681 LSSITVDLEKILKLTGEGDYSLRQTQTQFRDMTTAGKALLSKSNLRQGEKQKPLQBAQ 740
Db 551 LSSITV---LESTSSSNFNVNAEVANNSLASVNNSSSVLSSTADN-----LEINQ 601
QY 741 PFLGRVLDKKAIAKAKALVTKKATN-----GQLLERSINKAVLAYNNSAIKKANVK 793
Db 602 FGSDNLTKDSSEISTSGAFLSNQTSSSEASSNSMSSINSPLSLSLTSNSSESATNSQ-- 659
QY 794 RLEKELDLTLGLVEKGKPLAQATMVGQVYLLKTPLPYVYIGLVNVPYDKSGKLIYALDM 853
Db 660 -----SSEATKV-----DNNNS-----STHS 674
QY 854 SDTIGEGQKAVGNPILNVDENEGHALAVATLADYEGDLDTILNSKLQSLTSIROVP 913
Db 675 SNILNSGNDSDS---DSDSDSDS-----SNLSSSPNLETNQTISKPEVNNISENP 725
QY 914 TAAVHRAGIFQAIONAAABEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTN 967
Db 726 KKV-----SSNSVQENSTDEMSTNPKSS 750

RESULT 5
152300
gi|152300|ref|NM_001001001.1|giantin - human
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004
C:Accession: 152300
R:Shoda, M.; Misumi, Y.; Fujiwara, T.; Nishiohara, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A:Reference number: 152300; PMID:95100974; PMID:7802676
A:Accession: 152300
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
A:Cross-references: UNIPROT:Q14789; GB:D25542; NID:5662389; PIDN:BAA05025.1; PID:g808869
C:Superfamily: giantin

Query Match 4.2%; Score 214.5; DB 2; Length 3225;
Best Local Similarity 19.8%; Pred. No. 0.14;
Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;

QY 20 HNOEVFSLVKEPILKQTCASSISGADYAESGKSKLKI-----NETSGPVDDTVTL 72
Db 199 HEDELQLV-----TQAD-----VETEMQOKLRVLQKUBEHEESLVGRAQVVDL 243
QY 73 FSDKRTTPEKIDNLAQPREQELK-----AVTENTESEKQITSGSQLEQSKESLSLNK 126
Db 244 LQQLTAAEQRNQILSQIQWEAEHNTLVETERESKILLKMELEVAERKLSFH- 302
QY 127 TVPSTNSWEICDFTYKGNLTVGL-----SKSGVEKLSQTDHVLPSQAADGTQLIQV 178
Db 303 NLQEMHHLLEQFEQAGQAQAELESRYGALQKHKAEMEEKTSHLSLQKTG---QELQS 359
QY 179 ASFAFTPKKTAIAEYTSRAGENGFI-----SQLDVGDKKEIINEGEVFNLSYLLKKVTIPT 233
```

```
Db 360 ACDALKQONSKLLQDKNEQAVQSAQTIOQLLEDQLOQKSKEI-----SOFLNRLPQQ 411
QY 234 GYKHIGQDAFVDNKNIAEAVNLPESETISDYAFALHALKQIDLDPNLKAIGELAFPNQOI 293
Db 412 -HETASQTSPPDVN-----EGQAVTEENIASLQKRVVEL-----ENEK 450
QY 294 TGKLSLPRQLMRL-AERAFKSNHKTITFRGNSLKVTGEASFQDNDSLQMLPGLGLEKIE 352
Db 451 GALLSSIELEELKAENEKLSQITLLEAQRNT---GEA---DREYSEISIVDIANKRS 503
QY 353 SEAFPTGPNGDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDF 412
Db 504 SSA-----BESQD-----VLENTFSQKHKL---SVLLEMEAQBEI 539
QY 413 TYQKNSVTG-----PSNKGLOKVRKNKLE-----IPKOHNGVT 446
Db 540 AFLKLQLOKRAEADHEVLQDKEMQMBEGEAPIKMKVPLEDTGQDFPLMPNESSLP 599
QY 447 ITEIGDNAPRVDFONKTLRKYDLEE--VKLPSTIRKIGAFAPFOSNNLKSFEASDDLEEI 504
Db 600 AVE---KEQASTEHSQRTSEISLNDAGVELKST-----KODGDKLSLAVPDTGQC 647
QY 505 KEGAFMNNRIETLELKDVLKLTIGDAAPIHNIHVAIVLPESVQEIQR-----SAFRONGA 558
Db 648 HQDELERKLSQILEL-----ELNFHKAQEIYKKNLDEKAKEISNLNLQIEFFKNAD 699
QY 559 NN-----LIFMGSKVKTIGEMAFNLRLHLDL-----SBQKQLTEIPVQAQS 601
Db 700 NNSAFTALSEERDQLL---SQVKELSMVTELRAQVQKLEMLNLAERQERLDESOTAH 756
QY 602 DNALKEVL---LPASIKTIREBAFKQHLKQLEVASALSHIAFNALDDNDGDQFQDNK- 656
Db 757 DNLLTEQIHSLSIEAKSKDKVIEVL-QNELDDVQLQFSEOSTLIRSL-----QSLOQNK 810
QY 657 -----VVVKTHNSYALADGEHFTVDPDKLSSTIVDLEKILKLTGEGDYSLR 704
Db 811 SEVLGEARVNHISSKVEELSQALESQLELTKMDQLL-----LEK-----KDDVETLQ 859
QY 705 QTTQTPFRDMTTAGKALLSKSNLRQGEKQKPLQBAQFPLGRVLDKKAIAKAKALVTCKA 764
Db 860 QTIEBKQQVT-----BISPSMTEKMVLQNEEKFSLG-VET-KTLKEQLNLSRAE 909
QY 765 TKNQOLLERSINKAVLAYNNSAIKKA---NVKRLKELDLTLGLVEKGKPLAQATMVGQV 821
Db 910 AKGEQVEEDNEVSSGLKQNYDEMSPAGQISKEELQHEFDLLKKNQKQKRLQALINRK 969
QY 822 YLLKTLPLPPEYVYIGLVNVPYDKSGKLIYALDMSDT-IGEGOKDAYGNPILNVDDNDEGYH 880
Db 970 ELLQVRSLRBEELANLK---DESKK---EIPLSTERGEVEE-----DKNKEYS 1013
QY 881 ALAVATLADYEGDLTKILNSKLSQLSISR-----QVPTAAVHRAGI 922
Db 1014 EKVTSKQETIEIVLQKTISEKEVELQHIRKDLBEKLAABEQFQALVKQNNQTLQDKTNQ 1073
QY 923 FOATQNAAABEQLLPKPGTHSEKSSSES-ANSKDRGLQSNPKTNRGH 971
Db 1074 IDLLQAEISENQAIQKLTISNTDASDGSVALVKETVVISPPCTGSSEH 1123

RESULT 6
A56539
gi|A56539|ref|NM_001001001.1|giantin - human.
N:Alternate names: macrogolin
C:Species: Homo sapiens (man)
C>Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 09-Jul-2004
C:Accession: A56539; S37536
R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein (
A:Reference number: A56539; MUID:94187728; PMID:7511208
A:Accession: A56539
A:Molecule type: mRNA
```


Db 452 AKP-KKNSQGVGVKDKGLWYLYNESGSMATGWV-----KDKGLWYLYNESGSM-A 502

Qy 405 TKWLEED--FTYQKNS-----VTGF-SNKGL 427

Db 503 TGWVKDKGLWYLYNESGSMATGWVKDKGL 531

RESULT 8

C90538

hypothetical protein MYPV 2110 [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C:Accession: C90538

R:Chambaud, I.; Hellis, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: C90538

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-3216 <KUR>

A:Cross-references: UNIPROT:Q90Z9; GB:AL445566; PID:g14089624; PIDN:CAC13384.1; GSPDB:C90538

A:Experimental source: strain UAB CTIP

C:Genetics:

A:Gene: MYPV 2110

A:Genetic code: SQC3

Query Match 4.2%; Score 212.5; DB 2; Length 3216;

Best Local Similarity 19.9%; Pred. No. 0.17;

Matches 218; Conservative 148; Mismatches 380; Indels 347; Gaps 55;

Qy 7 TVALTLTT--VSVVTHNQEVPSLVK----EPILKQTOASSISGADYAEVSSGSKLKN 59

Db 2274 TLITITISSDOTSRTYQIEIGFVSDNRAENIVAKGYLSIEIOASGRAPESGTTVIQI- 2332

Qy 60 ETSGPVDP-----TVTDLFSDKRTTP-----KIKONLAKGPREQELKAVTENTSEKQITS 111

Db 2333 ----FILDPSVIONNSTDEFNKILTNHEASLGININKNPQ-----SNNSNLEQSILS 2381

Qy 112 QSOL----BQSKESSLNKTVPSTSN-----WEI-CDFITK-GNTLVGLSKSGVEK 156

Db 2382 AEVLTTNEVPAESVKLIQVSKKGDQDQDHYSVREVVVYSGFIQRTGADILNIRRAIEAS 2441

Qy 157 LSQTP----HLV-----LPSQADGTQLIQVASFAPDPK-KTAAEYTSRAGE 200

Db 2442 LRENDKAQHIVKVESREKITKIBESTTGVKNFFTTSDFEINNELIKTKVTNVQNIQDN 2501

Qy 201 NGE1-SQLDVGKEIINEGEVFNSSLAKKVITPTGYKHIGQDAFVDN---KNIAEVLNLP 256

Db 2502 NVELLSVLDSKRENIISGNARVKIRSK-----DNSSEKSEFIVNI-S 2544

Qy 257 SLETISDYAFALHALKQIDLDP-----NLKAIAGELAPP-----DNQITGKLSLPRL 303

Db 2545 GFETLSEAFDKYIKSSTRFIPSVKEANKDLMDTKATSENFPSQNKDNKFDL---- 2600

Qy 304 MRLAER-----AFKSNHIKTIEPRGSLKLVIGASFDQDNDLSQMLPDLGLEKTES 353

Db 2601 -AVKERFVFSVDTSVKPKFMELSKTIKTNFGEISTFFQSAQVEEDV-----DGFKK--- 2650

Qy 354 EAFQGNPGDDHNNRVVLWTKSGKNPSGLATENTVNPDKSLWQSPRIDYTKMLEEDPT 413

Db 2651 ELFDNT---DIFR---VQTEKDGK-----TINKYVKENE---KNSHNIDFTK----- 2688

Qy 414 YQKNSVTGFSNKGLOKVRKNKNLEIPKHQNGVTITEIGDGNAPRVNDFQNKTLRKYDLEEV 473

Db 2689 -----IKKLEI-----GNPSEI PKDYFKDAS-----SLTEL 2715

Qy 474 KLPSTIRKIGAPAFOSNNLKSFEASDDLEEEKGAFPMNRITETLEKDLVTIGDAAPHI 533

Db 2716 KINSQVSKIKESAFESAKLTSLLEPNSLVEIGPNAPKNSVLTSL----- 2759

Qy 534 NHIYAILVPESVQIEGRSAFRQNGANNLIF--MGSKVKTLCGMAFLSNRLEHLDLSEQKQ 591

Db 2760 -----SGLETKLSALKEN-----VFEKANDSIKTT-----IWNWKEALKAKEPI 2800

Qy 592 LTEIPVOAFSDNALKEVLLPASLTIRREAPKQNHKLQLEVASALSHIAFNALDDNDGDE 651

Db 2801 LT-----STNK-----IPSSD-----KKAKKPSQFEQ-----SNFSFAATKQEDSDI 2838

Qy 652 QPD-----NKVVVKTTHNSYALADGEHFIVDPDKLSSTIVDL-----EK 690

Db 2839 HFEITGFSQDDVFGTITIKIKIKTKTDESFFHVDEN-----NPSK---KVLTDFTKTNLEEK 2892

Qy 691 ILKILEGLDY-----STLRQTTOTQPRDMWTAGK---ALLSKSNLRQGEKQKFLQBAQF 741

Db 2893 VTKIKDASSVFDSSKIETIEKKRDTSSNNTSTENQNKYRLVQWKDVNKGSGSVFLPOSSV 2952

Qy 742 ---FLGRVDL-DKAIKAEKALVTKATKQNGQLLE--RSINKAVLAYNN----- 784

Db 2953 NAPPVETQIGDKTTSLVDFNNPQSQIRNNSLFENLKPSSAYLVMSNDASLKNQSKIT 3012

Qy 785 --SAIKANVKRLEKELDLTLGLVEGKPL-----AQATMVQGVYLL 824

Db 3013 LLSTFKNSSTHSLHFESKVLGSLWIPGSGSLRFDQNNINIDRESQDVNKANIESTENNITA 3072

Qy 825 KTLPLPLEYIG-----LVVYFDKSGKL-----IYAL 851

Db 3073 KGSWP-QSFIQGQESKTIILEYRDEQGRYEFTFFVFNNDKVKRYKVMKTEAVAKAYLI 3131

Qy 852 DMS---DTIGEGQKDAYGNPILVNDEDEGVALAVATLAD-----YEGLDIKTILNSK 902

Db 3132 DLSSIGDT-GNSKFOAKLSVISPNQNNFLTYEERIKLLDELGKKWFNNISIKDSSGK 3190

Qy 903 LSQLTSTRQVPTA 915

Db 3191 VENNTYNNMKTA 3203

RESULT 9

C97033

uncharacterized protein, probably surface-located [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C:Accession: C97033

R:Nolling, J.; Berton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: C97033

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1227 <KUR>

A:Cross-references: UNIPROT:Q97K41; GB:AE001437; PIDN:AAK79054.1; PID:g15023995; GSPDB:G97033

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1080

Query Match 4.1%; Score 209.5; DB 2; Length 1227;

Best Local Similarity 18.1%; Pred. No. 0.056;

Matches 237; Conservative 207; Mismatches 402; Indels 463; Gaps 68;

Qy 1 MKKHKLTVALTTLTVSVVTHNQEVPSLVKPEILKQTOASSISGADYAEVSSGSKLKLINE 60

Db 4 IKNTLLVPSISVTTSAV-----LLSKPAIKAYAADNSV-----YKSSISNSN-EINV 49

Qy 61 TSGPVDDTVTDLFSDKRTTPTEKIKDNLAKGPREQELKAVTENTSEKQITSGSLOESKE 120

Db 50 KGKEV-----QEKEVNVKNNDNSNSKVSSSNE-NQVSNKNSNPKVSSSSSEIQSINK 100

Qy 121 SLSL-----NKTVPSTGNWEICDEITKGT-----LVGLSKSGVEKLSQTDHLVLP 166

Db 101 NVNQLVQONNKSVLASGN--VDDVEKKINSQVTSQDYIAIGETKVKPDPDTLDIINKAIVD 158

Qy 167 SQAADGTQLIQVASFAPDPDKKTAIAEYTSRAGEN-GEISQLDVGDKKEINEGEV--FNS 223


```
Db 1649 -----NAVEGRKAKINAVTFSEYKDKALAKTEDA--- 1678
Qy 533 INHIYAVLPBSVOEIGBSAFQNGANNLIIPMGSKVTLGEMAFISNLEHL-----DL 586
Db 1679 -----YNAKVNEADN-----SNASTSSIEAKQKLAELKQTADQNV 1715
Qy 587 SEQKQLTEIPVQPSD-NALKEVLLPASLK-----TIREAPKK 624
Db 1716 NOATSKDDIEVQIHNLDNDINDYTIPTGKBSATTDLYAYADQKKNISADTNATQDEKQ 1775
Qy 625 NNLKQLE--VASALSHIAFNALDNDGDE-----QFNKVVVKTHNSYALA- 669
Db 1776 QAIKQVQNVQTALENIN-NGVNDGVDVDDALTOGKAADITQVATVPKANQAIEAKAE 1834
Qy 670 DGEHPIVDDKLSS--TIVDLEKILKLEGDYSTLRQTTQTFQFRDMMTAGKALLSKSNL 727
Db 1835 DTKEISIDHSQLTABEKEALAMIKITDQAKQGITDATTAEVEKAKAQGLEAFDNIQI 1894
Qy 728 RQGEKQKLOEAPFLGR-----VOLDKAIKAEKALVT-----KKATKNGQLLERSIN 776
Db 1895 DSTEKQKABELETALQDQIRAGVNVDDATTEEKEAFTNALEDILSKATED--ISDQFTN 1952
Qy 777 KAVLAYNNSAIKKANVKRL-----EKELDLTLGLVEGKGPLAOATMVQ 820
Db 1953 AEIATVKSALQKAEQINPVVKNALEATREVNVKQIEII-----KNADADASAKX- 2005
Qy 821 VYLLKTPLPPEYIYGLNVYFDK-SGKL-----IYAL-----DMS 854
Db 2006 --IART-----DLGRYFDRPADKLDKTQNTNEVAELQNVITPAIEAIVQNDPDAN 2054
Qy 855 DT-IGEGOKDAYGNPILNVDENEGHALAVATLADYEGLDIKILSKLSQLSIRQVP 913
Db 2055 DTNNGTDNDATANSANATPENTQOPNVSETT--DNGKADASPTTPNNSDAATGETTVT 2112
Qy 914 TAAVHRAGIFOAIONAAEAELQLP-----KPGTHSEKSSSES 952
Db 2113 SATDADKQKPOANNSSADASTNSTMDNDVTSKEPEVSTNGTIDKVPETDNATPAB 2172
Qy 953 A---NSKDRGLQSNPKTRNGRHSAILPRTGS 980
Db 2173 TTNNSTTTATNENAPTG---STATAPTAS 2200

RESULT 14
T31105
Hypothetical protein 2 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31105
R:Ward, C.K.; Lumley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, B.J.
J. Bacteriol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: Z20984; MUID:99030326; PMID:9811662
A:Accession: T31105
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4919 <WAR>
A:Cross-references: UNIPROT:Q9ZHL0; EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC7
C:Genetics:
A:Gene: lepA2

Query Match 4.0%; Score 204.5; DB 2; Length 4919;
Best Local Similarity 20.0%; Pred. NO. 0.75;
Matches 250; Conservative 181; Mismatches 415; Indels 405; Gaps 65;

Qy 10 LTLTVSVVTHNQEVFSIAVKEPIKQTAQ-----SSISGADYABSSGSKLK----- 57
Db 598 VTLSKSTLSAGELTFKKVKVNTLNNDSELANNLSNASHVNTLNKSKUSAKQADIK 657
Qy 58 -----INETSQPVDDTVTDLFSDKRTTPBEKI-----KDLN---- 87
Db 658 VNLTLNDTTELTAKMLDINSITITNNGTIAGIFAN--ITTEKLNKKEKALILAEQNLNFT 715
```

```
Qy 88 -----AKPREQELKAVTENTESEKQITSGSOLEBQSKESLSLN----- 125
Db 716 VNGSHYENKGDIVSKDKATVTFPSKNSDFTSGSKLVNAQNQLKVNNVNFTISQGGDITLI 775
Qy 126 -----KTVPSTSNWEICDPIFKGNTLVG-----LSKSGVE-----KLSQT 160
Db 776 GNVTLNASGFTNSGNLITVTKTLDVGDIQNFNTKGNLTUVGEDLHKSKTKITNDGKLISI 835
Qy 161 DHLVLPQQA---ADGTOL-IQVASFA---FTPDKKTAAIABYTSRAGENGESISQDV--- 209
Db 836 KNLNISSEADFINNGTTLGIEALKIATKGNFTNKEKAILASN-----SLLDISVA 885
Qy 210 DGRKEIINEGEVFNLSYLLKKVTIPFG-----YKHQDAPFVDNKNIAEVL----- 254
Db 886 EGKKTFFNG-----TIESGNLNTINTGAFNLVDNATIRSFVLNITSTGNVSN 934
Qy 255 -----PESLETISDYAPAHALAKQIDLPD-----NLKAIGE 285
Db 935 NGTILSNERLNITSAANFTNESGNTVMSNGLLNIIAQGNITKNKLIASRQQLNLTAVAD 994
Qy 286 LAPFDNOITGKLSLPROLMRLA---ERAPKSNHNIKTIEFRGNSLKVIGESAFQDNLSQ 341
Db 995 NITNDSNISNKIAVLHSLGNISLSKQVYNLGEI---YAGNNISV--KAHQLKNDV-K 1047
Qy 342 LMLPDGLEKIESEAF---TGNPGDDHYANRVVLWTSGKNPSGLATENTYVNPDKSLW 396
Db 1048 LMGDITTTKKGQASYKLYQASNGG--HFGND-----GSSGYSEGDNLKKGKPADLNDKLT 1101
Qy 397 -QESPEIDYTKWLEEDFTYOKNSVTGFS---NKGLOQVKRNKNLEIPQHNGVITIEGD 452
Db 1102 VQRTGKI---YAGRDUTFNKSNAGKSEIINRGTINVKKLSYD-----SD 1144
Qy 453 NAFEN-----VDFQNKTLR-KYDLEEVKLPSTIRKIGA---PAPOS--NNLKSEASDD 500
Db 1145 VSPENMQSKVDLYTKIPEAKSDIE---LTFKTNTHPVYLVFNKNNNEKKYRNSEN 1199
Qy 501 LEEIKGAFMNNRIETLEBKDKLVTDGAAPHINHIYAVLPESVQVIGISAFQNGANN 560
Db 1200 TKNFK-----SIGDL--INEALSDSAPEAIE---AYTSGSSN 1232
Qy 561 LIPMGSKVKTLGEM-----AFLSNRLEHL-----DLSEKQLTEIPVQAPSD----- 602
Db 1233 YINPVSVYLAALGNANSSNPHYLNLTALKHILGNQWQDLDKKQENI-KVLQKWEDEPKOK 1291
Qy 603 NALKEV-LLP-----ASLKTIREEAFKKNHLKQLEVASALSASHIAPNA 643
Db 1292 GASKMLDLYPNTDKEKAKIFAGIIRNGNDTISDVESEDFKKYKSKFQNGEAKNDTGTDS 1351
Qy 644 LDDNDGDEQFNDKVVVTHH---NSYALADGEHFIV-----DPDKLSS 683
Db 1352 YDSTKASEKY--KKVENVDHKNIDEHKLNIKGHEITVPGVSPENLNKNMNDHQDPKLGE 1409
Qy 684 TIVDLEKILKLEGDYSTLRQTTQTFQFRDMMTAGKAL--LSKSNLRQGEKQKFLQEAQF 741
Db 1410 --IDKSIISELLAQPVVTEKSAARDSPRVNDNKEALDNLYRTRLSYINQNYL-GAKY 1466
Qy 742 FLGRVLDLKAIAKAEKALVTKKATKN---QQLERSINKAVLAYNNSAIKK-----ANVK 793
Db 1467 PFNQLDTE-----DKLKGIRIGDNYFEHQILTRLTEK--VADNHLTLKHLHDIALVK 1519
Qy 794 RL-----BKELDLTG---LVGKGPLAQ-----ATWVGVLKTPPLPEYIIGL 837
Db 1520 KLIDSASIQAKDLMLKVEALTKEQKDNLKEDIWVYVKTVEVNGQEVL-----VPQVYL-- 1572
Qy 838 NVYFDKSGKLIYALDMSDTTIGEGOKDAYGNPILNVDS-----NEGYHALAVATLADYE-GL 893
Db 1573 -----AKQIEEVEKORGVTGQIRA---GIIDVKVDVDRNTG-----TIAGYAVGL 1616
Qy 894 DIKT-----ILNSKLSQLTSIRQVPTAAVYHRAGIFQAIQNAAAEABQLPKPGTH--- 943
Db 1617 EAKNKLKNTGDILSRLSKL-----VGKKGLESTGVTVVDGTATKVKARIKSEGHIYL 1671
```


Db 1087 LDANSVDDTKAAIKESDAMOF---DSVQDLNGDIFNN-----TKDQVAPLND----- 1132
Qy 534 NHIYAIYVLPESQVQIEGRSFRONGANNLIIFMGSKVITLGEAFPLSNRLEHLDLSEQKOLT 593
Db 1133 -----LLEKAE-OKS-ISAENANTLI---OKDELAQAISIEGVVVKINRDEVIKOR 1180
Qy 594 EIPVQAFSD-----NALK-----EVLPLASLKTREE----- 620
Db 1181 KVKLDAYNDMVTYSNKLMTKEVNNAIKTLNADTLRIDSLKLKRLKLDMSBAELSDLEV 1240
Qy 621 -----AFKKNHLKOLE-----VASAL-SHIAFNALDDNDGDEQPD 654
Db 1241 KSINNVDADAKELKLEEKMLQPGGYSNSQIEAMQSVKSALESYI--SASEEATSTQEMN 1298
Qy 655 NKVVVTHNSVALADGEHFIVDPPDKLSSTIVD-----LEKILKLEGL-----DYSTLR 704
Db 1299 KQALVEAGTSLNWTDOQKANEETKTSYVVDKYKEALEKVNAEIDKYNQVNDYPKYS 1358
Qy 705 OTTOTQFRD-MTTAGKALLSKNLRQGEKQKELQEAQFFLGRVLDLDAKAEKALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQKKLMQ-EQAKLLKD-QIKSGNI-----TOYGLVITST 1404
Qy 764 ATWNG--QLLERSINKAVLAYNNSAIKANKVRLEKELDLTLGLVEGK-GPLAQATWVG 820
Db 1405 TSSGGTSPSTGSGYSKYSYNSAASKYNV-----DPALIAAVIQESGFNAKARSVG 1459
Qy 821 VYLLKTPPLPEYIYGLNVVFPKSGKLIYALDMSDTIGSGOK-DAYGNPILNVDEDN 879
Db 1460 AMGLQMLPATKSLGNNAYDP-----YONVGGTKYLAQLEKFGG--NVEKALAA 1511
Qy 880 HALAVATLADYEGLD-----IKTIL---NSKLSQLTISRQVPTAAVHRAGIFQAIQ 928
Db 1512 NA-GPGNVIKYGIIPPEKTONVVKIMANYSKLSSTAS-----SIASY----- 1556
Qy 929 AABAEQLLPKPGTISEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG--SFVY 986
Db 1557 -----TNSAPRVSSKYQGESGLRSSP--HKGDTFAAKAGTAIKSLQSGKV 1601
Qy 987 GILGYTSA 995
Db 1602 QIAGYSKTA 1610

RESULT 18
H89960
hypothetical protein SA1577 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: H89960
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2186 <KUR>
A:Cross-references: UNIPROT:Q99TB0; GB:BA000018; PID:g13701552; PIDN:BA842845.1; GSPDB:G
C:Experimental source: strain N315
C:Genetics:
A:Gene: SA1577

Query Match 3.9%; Score 199; DB 2; Length 2186;
Best Local Similarity 19.2%; Pred. No. 0.4;
Matches 217; Conservative 182; Mismatches 387; Indels 342; Gaps 58;
Qy 20 HNOVFSLVKEPILKQTOASSISGADYABSSGSKLKNETSGPVDVTLDFSD---- 75
Db 1208 YNAKLAENATPDATNDEKNAINTLNQDROQAIESIKQANTNAEVDQAATVAENNIDAV 1267

Qy 76 -----KRTTPKIKDNLAQPREQELKAVTENTSEKQITSGSQESKE-----SLSLN 125
Db 1268 QVDVVKQQAARDKITAFAVAK--RIEAVQTNPATDEEKQ-AAVQINQLKQQAQINQIN 1324
Qy 126 KT---VPSTSNWEI-----CDFITKGNLTVLGSKSGVEKLSQTDHLVLPQAAAGTOL 175
Db 1325 QTNQDQVDTTQAVNAIDNVEAEVVKRAIADIEKAVKEKQQQIDNSL---DSTDNEK- 1380
Qy 176 IQVASFAPTPDKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFNLSLLKAVT--IPT 233
Db 1381 -EVASQALAKEKEKALAA-IDQAQNSQVNOQAATNGVSAIK-----IIQETKVKFA 1430
Qy 234 GYKIGQDAFVDNKNIAEVLNPE-----SLETISDYAFALAL-----KQID 275
Db 1431 AREKINQKA---NELRAKINQDKBATABERQVALDKINE--FVNQAMTDTITNRTNQVD 1485
Qy 276 LPDNLKALGELAFPNQITGKLSLPQLMRLAER-----AFKSNHITITIFRGNLSKV 328
Db 1486 -DTTSQALDSIA-----LVAPETHVRAAADAVKQVBAKQIEQAEASHATDEEQ 1535
Qy 329 IGEASFQND-----LSQMLPDGLEKTESAFNGPDGDDHNNRVNVLWTKSGNPSGL 382
Db 1536 VALNQLANNEKALQINQINQAVTNDVKKVETNGIATLKG---VQPHIVI-----KPEAQ 1587
Qy 383 ATENTYVNPDKSLMQESP-----EIDYTKWL-----EEDFTYQKNSVTGFSN--- 424
Db 1588 AIKATAENQVESI-KDTPHATVDELDEANQLISDTLKQAQOEIENTNQDAVTDVNRQTI 1646
Qy 425 KGLQKVK-----RNKN-----LEIPKQHNGVTITEIGD--NAFRNVD 459
Db 1647 KAIEQIKPKVRRKRAALDSIENNNKQDLAIRNTLDTTQDERDVAIDTLNKIVTINKDI 1706
Qy 460 FQNTLRKYVLEE-----VKL-----PSTIRKIGAFQSNLKSFEASD-DLEBI 504
Db 1707 AQNKNAEVRTETDGNNDNIKVLKPVQVKAAROSVGVKBAQNAL--IDOSDLSTEE 1764
Qy 505 KEGA-----FMNNRIETLEKDKLVTIGDAAFINHIYAIYLP-----ESVQEIG 549
Db 1765 RLAAKHLVEQALNQAIQDINHADTKAQNODSIDAQNIISKIKPATTVKATALQIQNIA 1824
Qy 550 RSAPFQNGANN-----LIPGSKVK--TLGEMAFPL-----SNRLEHL 585
Db 1825 TNKINLIKANNEATDEEQNIAIAQVEKELIKAKQOIASAVTNADVAYLLHDEKNEIRETE 1884
Qy 586 -----LSEKQLEITEIPVQAFSDNALKEVLPLASLKTIREAFKQHLKOLEVASALSHI 639
Db 1885 PVINKKASARQITTL-----FND---KKQAIENAIQATVEE--RNSILAQLQ---NIYDT 1932
Qy 640 AFNALDDNDGDEQFND--KVVVVTHNSYALADGEHFIVDPPDKLSSTIVDLKILKLE 697
Db 1933 AIGQIDQDRSNAQVDKTSASLNQTIHDL-----DVHPIKKPDAEKTINDDLARVTALVQ- 1986
Qy 698 LDYSTLRQTTQTFQDMRTTAGKALLSKNLRQGEKQKFLQBAQFPLGRVLDLDAKAEK 757
Db 1987 -----NVRKVSNNRK-----ADALK 2001
Qy 758 ALVTKKATNGQLLERSIN---KAVLAYNNSAIKANKVRLEKELDLTLGLVEGKPLAQ 814
Db 2002 AITALKQMDBELKTPARTNADVDVLRKFNVALSDIEAVITEKENSLLR--IDN---IAQ 2056
Qy 815 ATMVGQVYLLKTPLPPEYIYGLNVVFPKSGKLIYALDMSDTIGSGQKDAVGNPILNVDE 874
Db 2057 QT-----YAKPKATATPEQLAKVKVLDIQ-----YVAD-----GNRM--IDE 2091
Qy 875 DNEGVALAVATLADYEGLDIKTILNSKLSQLTISRQVPTAAVHRAGIFQAIQNAABAE 934
Db 2092 D-----ATLN-----DIKQHTQFIVDEILAIAK-LPAEA-----TKVSPKEI 2126
Qy 935 QLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG 982
Db 2127 QPAPKVCTPIKKEETHESRKVEKE-----LPNTGSEG 2158

Query Match 3.9%; Score 197.5; DB 2; Length 1790;
Best Local Similarity 21.0%; Pred. No. 0.35;
Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

QY 25 FSLVPEPILKQQA---SSSISGADYABSSGSKLKNITSGPVDVTP-----71
DB 434 FLQORDFLKQVLCNNSTNNVGDNAKENGSKSDKSDKDTGDKGTGYESFKAN 493
QY 72 LF-----SDKRTTPEKI-----KDLAKGPREQELKAVTENTSEKQITSG 112
DB 494 LFEVLLNYDALNLPKLFPTTDFPMFFQDQ-----KYSELREITRN-----VTTG 543
QY 113 SOLROSKESLSLNTVP---STS-----NWEICDPTIKGNTLVGLSKS 152
DB 544 NDLE-DEEPLKAIQTISELLTSLTAADIRIPISYLTFLIYWLFGDFKATNDFL--SDKS 600
QY 153 GYVKLSQTDHLVPSQAAD-----GTQLQIVA---SFAFTPDKTAIAEY--TSRAGEN- 201
DB 601 VIKSLLSFSYQI---QDEVDTIKCLVTMLGVAYEFSSKESPPFRKEYFEFTIKLGDN 657
QY 202 --DEISQDLDV---GKEIINEGEV-----ENSYLLKKVT-----IPTGVKH 237
DB 658 YASRIKQPKDSYFQVNDNEISILTPDELGLPKVYFSTFIQLFNENIYRIITLASH 717
QY 238 IQGDAFVDNKNIAEVLNPE-----SLETISDYAFALHAKQIDLPNKLKAIGEL 286
DB 718 DPDEEPINKISPEEVEKLQROCTKLGEITSLQTESHTENLTKLIALTNEHRELDEK 777
QY 287 AFPDN-----QITGKLSLPRMLRAEPKSNHIKTIIFRGNLSKV 328
DB 778 YOILANSSHSLSKENFSILETELKNVRDSDLEMTQLRDVLETKENQOTALLEYKSTIHK- 836
QY 329 IGHASFQDNDLSQMLPDGLKIESAFTGNPDHYNRVLVLTSGNQPGLATENTY 388
DB 837 -----QEDSIK--TLERGLETLISQK---KKAEDGIN-----KMGKOLFALSRMOQA 878
QY 389 VNPD--KSLMQESPEIDY-----TKMLEEDFTYQKNSVTGFSNGKLGQVK--RNKNLEIPKQ 441
DB 879 VEENCKNLQEKDKDNNVNHQETKSLKEDIKAIKITEIKAI--NENLEEMKIQCNNLSKEKE 937
QY 442 HNGVITITEGD-----NAFRNVPQNTLRYKIDLEEVKLPSSTR- 480
DB 938 HISKELVEYKSRFQSHDNLVAKLTKLSLANNYKDMQAEENSLIK-AVEESKNESLIQL 996
QY 481 -----KIGAFQSNLKK-----SFEAS-----DDEEIKEGAFMNNRIETLE-----518
DB 997 SNLQNKIDSMQKENFQIERSIEKNIIEQLKTTISDLEQTKBEETISKSDSKDEYESQI 1056
QY 519 --LQDKLVITIGDA-AFPHINHIYAIVLPSVQBIGRSAPRQNGANNL-IPMGSKVKTGLGEM 574
DB 1057 SLKKEKLETATTANDENVNKNISLTKTRELEALEAAYK-NLKNLETKLETSEKALKEV 1115
QY 575 AFISNRLEHLDSEQKQLTEIPVQPSDNALKEVL-----LPASLKTIREFAFKKNHLK 628
DB 1116 KENEELHKEEIKLEKEATEQQLNSLRANLESKEHEDLAAQLKYEEQIANKERQY 1175
QY 629 QLEVASALSHPALNDNDGDEQPDNVKVVKTHNSYALADGEHFIV-----676
DB 1176 NEEISQ-----LNDEITQTQENESIKKKNDE-----LEGEVKAMKSTSEOSNLKK 1222
QY 677 -DPDKLSSTIVLEK-----ILKLEGLDYSL-----ROTTOTQFRDMTT 716
DB 1223 SEIDALNLOIKELKKNETNEASLLESIKSVSEVTKIKELQDECNPFKEVSELEDKLL 1282
QY 717 AGALLSKSNLRQGEKQKFLQ-----BAQFPLGRV-DLDRKAIKAEKALYTKKAT---765
DB 1283 ASQDKNSKYLELQKESKEIKELDAKTTELKIQLEKITNLKSAKESSELSRLKTSSE 1342
QY 766 --KNG-QLLERSINKAVLAYNNSAIKANVKLEKELDLTLGLVEGKGLPQAATVQGVY 822
DB 1343 ERKNAEQLKLEKLN---IQIKNOA-----FEKERKLLN---EGSSITITO-----1381

QY 823 LKLTPLPEYVYIGLVNVPFDSKGL-----IYALMSDTIGEQKXDYGNPILNVEDNE 877
DB 1382 -----EYSEKINTLEDELIRLQENELKAKEIDNTRSELRKSVLSNDELLEKQN- 1431
QY 878 GYHALAVATLADYEGDLIKTILNSKLSQITSI-----ROVPTAAYHRAGIFQAIQNAAE 932
DB 1432 -----TIKSLQD-EILSYKDKITRDEKLLSIEDNRKDLSEKQD-----LRAAQESKAK 1481
QY 933 AEQLLPKGTTHSEKSSSESA---NSKD--RGLQSNPKTN 967
DB 1482 VEEGLKK---LEESSEKKALEKSKEMMKKLESTIESN 1517

RESULT 21
B70126
surface-located membrane protein 1 (lmp1) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 16-Aug-2004
C:Accession: B70126
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kervage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: B70126
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1119 <LE>
A:Cross-references: UNIPROT:O51228; GB:AE001131; GB:AE000783; NID:g2688098; PIDN:AAC6659
A:Experimental source: strain B31
C:Superfamily: tetratricopeptide repeat homolog
F:742-774/Domain: tetratricopeptide repeat homolog
F:775-808/Domain: tetratricopeptide repeat homolog <TT2>
F:809-842/Domain: tetratricopeptide repeat homolog <TT3>
F:843-876/Domain: tetratricopeptide repeat homolog <TT4>
F:911-943/Domain: tetratricopeptide repeat homolog <TT5>
F:944-977/Domain: tetratricopeptide repeat homolog <TT6>
F:979-1010/Domain: tetratricopeptide repeat homolog <TT7>
F:1011-1044/Domain: tetratricopeptide repeat homolog <TT8>

Query Match 3.9%; Score 197; DB 2; Length 1119;
Best Local Similarity 19.6%; Pred. No. 0.18;
Matches 212; Conservative 164; Mismatches 358; Indels 347; Gaps 54;

QY 18 VTHQVPSLVKEPILKQTAQSSISGADYABSSGSKLKNITSGPVDVTPDLSKCR 77
DB 188 VNDQKNLFLEK---LKK-----NLGSKNSSENTILNDSQKIE-----NDKQ 225
QY 78 TT---PEKIDNLAGPREQ-----LKAVTENTSEKQITSGSQLEQSK---ES 121
DB 226 NTNLSKESNENILKTPDNSKYSNNNTSLKSSNSQKSELSPPSQTTIGKIYRPVS 285
QY 122 LSLAKTVPSTNWEICDPTIKGNTLVG-----LSKSGV-EKLSQTDHLVPSQAADGTQ 174
DB 286 YLIKEL-----YEILDDINTGRVTLGKNRLKELIKGSLNKFQKVNLIENSKNEASN 340
QY 175 LI-----QVASFAPTPDKKTAIAEYTSRAGENGESIQDLDVGEKEINEGEVFNYSY 224
DB 341 LLLTLIKKIDBPNLINIPKPYK-----EIFOLDKEDKK-----POY 378
QY 225 L-----LKKVTIPTGYKHIGQDAFVD-----NKNIAEVLNPELSTETI 261
DB 379 LEDLKSQVHSIKPIDLENTKSR--QQAIFKDLNEFLKNNPNDQAASKTLAQNKIHLBDL 436
QY 262 SDYAFALHAKQIDLPNKL---AIGELAPF-----DNOITGKLSLPRQLMRLAERAF 311
DB 437 KSKVH---SIRPIDL-ENTKSRQQAIFKDLNEFLKNNPNDQAASKTLAQNKIHLLEDLS 492
QY 312 KSNHIKTIET-----RGNLSKVIKEASQDND-----LSQMLPDGLEKIESEAPT 357
DB 493 KVHSIKPIDLENTKSRQQAIFKDLNE--FLKNNPNDQAASKTLAQNKIHLLEDLSKSVHS 550

Query Match	3.9%	Score 196.5;	DB 2;	Length 1875;
Best Local Similarity	20.8%;	Pred. No. 0.42;		
Matches 242;	Conservative 164;	Mismatches 378;	Indels 379;	Gaps 59;

Qy	34	KOTQASSSI-SGADYAYESSKSKLKINET----	-SGPVDVTITOL-----	FSDKRT	78
	:	:: :	: ::	:	:
Dd	199	RKTQEHLTQSNDNWLEKELRSK--NEQYLSVROKTDKVILDIRNELRLRDNDFOWERT	255		
Qy	79	TPEKIK---DNLAKGPREQ-ELKAVTENTESKQITSG-----	-SOLEFSQS	120	
	:	:: :	: ::	:	:
Dd	256	NNDLVKQKNELSKSLQEKLLRIKGLSDLSNSKQFBSAEMSLKQRVLVDLLESQLNAVKE	315		
Qy	121	SLS----LN--KTVPSTSNWEICDPFITKNTLVGLSKSGVEKLSSOTDHLVLPQA-----	169		
	:	: :	: :	: :	: :
Dd	316	ELMSIRELNTAKVIADDSKKQ-----TPENEDLLKBLQLTKEKAQOCEKCLRUSSTIDEA	371		
Qy	170	-ADGTOLIQVASFAPTDPKKTAIAETYSRAGENGEISQLDVDGKE--IIN-----	216		

Db 372 DEDNENLAKSSDFIFKKQLIKERRTKHEHQNOIFTFIVELEHKVPIINSFKERTDML 431
Qy 217 EGEVNSYLLKKVTPTGYKHIGQDAFVNDKNIAEVNLPESLETISDYAFALHAKQIDL 276
Db 432 ENELNNAALL-----LEHTSNE--KNAKVKELNAKQ-----KLVEBC 466
Qy 277 PDNLKAIGELAFFDQITGKLSLPQLMRLABRAFKSNHIK-----TIEP----- 321
Db 467 ENDLQTLTK-----QRDLRCIQIYLLITNSVSDSGPLRKEBQIPIQIMQBD 517
Qy 322 ----RGNSLKVTIGASFOONDSQLM-----LPGLE-----KISE 354
Db 518 STITESDQKVTELVFKNIIQOEKNAELLKVVRLNADKLESKEKSKOSLOKISE 577
Qy 355 AFTGNPGDDHNNRVVLWTKSGKSPGLATENTYVNDKSLMQWESPEIDYTKWLEEDFTY 414
Db 578 --TVNEAKE-----AIITLKSEK-----MDLESRIEELQKELEELKTS 613
Qy 415 QKNSVTGFSNGLQKVRNK-NLEIPKHQNGVTITEIGDNAFRVDFONKTLRYKDYDEV 473
Db 614 VPNEADASYNTIKQLTGKRDLESQVQDLOTRISQITRESTENMSLLNKEIQ--DLYDS 671
Qy 474 KLPTIRKIG-----AFAPQSNLKSFRASDDLEIEKEGAFNNRI----- 514
Db 672 KSDISI-KLGKESSRILAEERFKLLSNTLDLTAKEND-OLRKRPDYQLONTILKQDSKTH 729
Qy 515 ----ETLEKOKLVITIGDAAPHI--NHIYAIVLPEV--QEIGRSAPRQNGANNLIFMGSK 567
Db 730 ETLNEVSVCKSLIVETELLNKEEQKLRVHLEKNLQELNKLSPK---DSLRIWYQ 786
Qy 568 VKTIGEMAFNLRLHLDSEQKQTEIPVQAFSONALKEVLLPASLTIRBEAFKQHL 627
Db 787 LQTLQKER--EDLLEBTRKSCQKIDEL-----EDALSE-----LK--KETSQKH 830
Qy 628 KOLEVASALSHIAFNALDNDGDEOFDNKV--VVKTHNSYALADGEHFIVDPKLSSTI 685
Db 831 KOLEP-----DNNSIEWQNKIEALKKDYESTVSDSQ--TDIEKLQYKV 876
Qy 686 VDLEK-----ILKLEIG-----LDYSTLRQTT--- 707
Db 877 KSLKEIEBEDKIRLHTYVMDETINDSLRKELEKSKINLTDAVSQIKEYKDLVETTSQS 936
Qy 708 --QT-----QPRDMTTAGKALL--SKSNLRQEGKQKFLQBAQFFL-GRVLDLKAIAKAEK 757
Db 937 LQOTNSKLDSEGFKQFTNQIKNLTDEKTSLE--DKISLLKBOQFNLNNELDLQKGMKEK 994
Qy 758 ALVTKKATKNGQLLERSINKAVLAVNNSAIK-----KANVKRLEKELDLTLGLVEGKGP 811
Db 995 ADFKURI-----SILQNNKEVEAVKSEYKSKIQNDLDOQT----- 1033
Qy 812 LAQATMVGQVYLLKTPPLPEYIIGLVYFVKSGKLIYALDMSDTIGE--GQKDAYGNPI 869
Db 1034 -IYANTAQNNY-----EQELQKHADVSKTISELREQHTYKQV 1071
Qy 870 --LAVDED-----NEGYHALAVATLADVEGLDITILNLSKLSQLSIRQVPTAATHR 919
Db 1072 KTLNLSRQLENALKENKSSQKESL--LEQLDLS--NSRIEDLSQNKL--LYDQ 1123
Qy 920 AGIFQAIQNAAAEQQLPKPQTHS-----EKSSSES 952
Db 1124 IQIYTA---ADKEVNNSTNGPOLNILITLRRERDILTQVTVABERDAKMLRQKISLMDV 1180
Qy 953 ANSKDRGLQSNPKTWGRHSAIL 975
Db 1181 ELQDARTKLDNSRVEKENHSSII 1203

RESULT 25

T18296

myosin heavy chain - Entamoeba histolytica

C:Species: Entamoeba histolytica

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18296

R:Guillen, N.

submitted to the EMBL Data Library, February 1997

A:Reference number: Z18865

A:Accession: T18296

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2139 <GUI>

A:Cross-references: UNIPROT:Q07569; EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB480

C:Genetics:

A:Gene: mhca

C:Superfamily: myosin heavy chain; myosin motor domain homology

F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 3.9%; Score 196.5; DB 2; Length 2139;
Best Local Similarity 21.7%; Pred. No. 0.51;
Matches 228; Conservative 121; Mismatches 369; Indels 331; Gaps 52;

Qy 1 MKKHLKTVALTITVSVVTHNQVPSLVKEPILKQTAQSSISGADYAESGSK---LK 57
Db 1237 IEKEMKALQEBKENVESKNSKTEKDKKLELDNLDKDTQKLDMDTADNEKAKAKDLEAQ 1296
Qy 58 INETSGPVDDTVTD--LPSDKRTTPEKIKONLAKGPREQLKAVT-----ENTES 105
Db 1297 LNEVDNHERKAVADAELLNKKKAQSDKELNSL-----KAELEALTAKSVVESKNKUSEN 1351
Qy 106 EKQITSSQLESQESKESLANKTVPSTNWEICDPITKGNLTVLGSKSGVEKLSQTDHLVL 165
Db 1352 EKAALS-EEIDQANEKL---KNIQADLRKATAD-LQEBANE---KKAEBE--AQORDKLVA 1400
Qy 166 PS-----QAADGTQLIQVASPAFTPKKTAIAEYTSRAGENGESOLDVDGKEII 215
Db 1401 DNKQMTKLEIKARDEENTYKVENYKVLKRGKADLE---EANEEN-----LDIEKKDRM 1452
Qy 216 N-----EGEV-----FNSYLLKKVTIPTGVKHIGQDAFVNDKNIAEVNLPESLETI 261
Db 1453 NKEQVKKLEGELEKTKDKLNAAIAEKOSIFTAKKQSDADLELNKTVEE----- 1502
Qy 262 SDYAPAHALAKQIDLDPNLKAIGELAFDQITGKLSLPQLMRLABRAFKSNHIKTIEP 321
Db 1503 HDEVVAKLNTQITKLTTR-----DNQ-----SABEELNELRSKADKDK----- 1539
Qy 322 RGNLSKVTIGRASFOONDSQLMLPDGLEKTESBFTGNPGDDHNNRVVLWTKSGKPSG 381
Db 1540 ----KKISELEGEQVNELES--RPVG-----TGN----- 1561
Qy 382 LATENTYVNDPKSLQWESPEIDYTKWLEEDFTYQKNSVTGFSNGLQKVRKNKLEIPKQ 441
Db 1562 -ADENEIKIRDAQI-----ADLNKALE-----MKGVQNNQLQAT---NKELKAK 1601
Qy 442 HNGVTI--TEIGDNAFRNVDFONKTL-----RKVDLEEVKLPSTIRKIG 483
Db 1602 DNDLTSKIEITENEMKKLENKAKRLEQDKQADKAVSEQTIKRGLEBEVVKLT--BIQ 1659
Qy 484 AFAPQSNLKSFEASD-----DLEIKSGAFNNRIETLEKDKLVITIGDAAPHINHI 536
Db 1660 ALKFOINAPSVAQEEKQRLSEDIABEKE-----QLEOERTTAANAERKKI 1708
Qy 537 YAIVLPSVQIBGRSAFRQNGANNLIFMGSKVKTIGEMAFLSNLEHLDLSEQQLTIEP 596
Db 1709 QA-----ELDEVKNLEDTVN-----QREKLVAKNS--ENDAEIDSLKEEKKALEDE 1753
Qy 597 VQAFSD--NALKE-----VLLPA-----SLKTIRE--EAPKKNHLK- 628
Db 1754 IEKITDDNNKLESEIDSLDRKYNALLDSKQSDVSMKSKFQDELKVTQDALETEKNHAET 1813
Qy 629 -----QLEVASALSHIAFNALDNDGDEQDPKNVVKVTHNSYALADGE-----HFI 675
Db 1814 MRLKGRLEKAAEVQVRLLEALQKNLDLAQKEKAKATK-----DYRAADGELKSLNLELDDV 1869
Qy 676 VDP--DKLSSTITVDLEKILKLTJEGLDYSTL-----RQTTQTPRDMTTAGKALISK 724
Db 1870 KQQLDKAQDLDADKDEDELATLD-QKYKTLVKQSVFDSRIQEMQSOL--DLEKAGRAKAKQ 1927

Qy	725	SNLRQGEKQKFLQEAQFGLGRVLDLDAIAKAEKALVTKKATNGQLLERSINK-----	777
		: : : : : : : : : : : : : : : : : :	
Db	1928	QKQAYEKKLOELQE-----NDNDFEYKATDKRINTLSAQKDD--LQKELEKERGLQKD	1980
Qy	778	-----AVLAYNNSAIKKANVKRLKELDLDTG-----LVEG	808
		: : : : : : : : : : : : : : : : : :	
Db	1981	SEKEVQLRVKQCELETKVAEFGVGANYSIAKVKAK-YEAEIEELTTAEADLAKAKMAEK	2039
Qy	809	KGPLAQATMQGVVLLKTPLPPEYYITGLNVYFDKSGKLIAL-----DM	853
		: : : : : : : : : : : : : : : : : :	
Db	2040	KAKTSQKKLDE--LQKT---IADYETKEASFNTIEGKTQAELEKKYQQQVRDDETRMSSL	2093
Qy	854	SDTIGEGQKDAYGNPIILNVDEDNEGYHAL	882
		: : : : : : : : : : : : : : : : : :	
Db	2094	EDEIKKG-TDALANKQLELDKVNKQYEKL	2121

RESULT 26

AF1116

internalin proteins, probable peptidoglycan bound protein (LPXTG motif) homolog lmo0333

C/Species: *Listeria monocytogenes*

C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C/Accession: AF1116

R/Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blocker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H. D.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001

A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Makok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A./Title: Comparative genomics of *Listeria species*.

A/Reference number: AB1077; MUID:21537279; PMID:11679669

A/Accession: AF1116

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1778 <LA>

A/Cross-references: UNIPROT:Q8VA32; GB:NC_003210; PIDN:CAC98412.1; PID:g16409711; GSPDB: A/Experimental source: strain EGD-e

C/Genetics:

A/Gene: lmo0333

```

Query Match      3.8%; Score 195.5; DB 2; Length 1778;
Best Local Similarity 20.2%; Pred. No. 0.43;
Matches 209; Conservative 137; Mismatches 330; Indels 361; Gaps 50;

QY 44 GADYAESSGKSLKINETSFPVDVTVDLSPDKRTT--PEKIKD--NLAKPQROELKAVT 100
    ||| : : : : : ||| : : : : : ||| : : : : :
DQ 27 GADETTVSEDTAVKTAEDASGATESGDSDDTEAPEKPEAKEASAKETTEKEEKAKT 86
    ||| : : : : : ||| : : : : : ||| : : : : :
QY 101 ENTESEKQI---TSGSQLEQSKESLSINKTVTPSTSNWEICDFITKGNLV---GLSKSGV 154
    ||| : : : : : ||| : : : : : ||| : : : : :
DQ 87 EEPASNKTREINTDKSQLKT---SLKAAVPAGSTY-----NSLPFDDNLAK--- 130
    ||| : : : : : ||| : : : : : ||| : : : : :
QY 155 EKLSQTDHLVLPQAAADGTQLIQVSPAFPTDKKTAIAEYTSRAGENGESQLDDVGKEI 214
    ||| : : : : : ||| : : : : : ||| : : : : :
DQ 131 -KLAV---IITGNAATGNSVDSAALL-----AISQLDL----- 161
    ||| : : : : : ||| : : : : : ||| : : : : :
QY 215 INGEVPSYLLKKVTPPTGYKHIGQAFVDNKNIAEVNLPESLETISDVAFHALLKQI 274
    ||| : : : : : ||| : : : : : ||| : : : : :
DQ 162 --SGETGND-----PTDISNIEGLQYLE--NLTSNLSEN--NISDLA---PLK-- 201
    ||| : : : : : ||| : : : : : ||| : : : : :
QY 275 DLPDLNKAIGELAFFDNOITGKLSLPQLMELAEAPKSNHIKTIE--PRGNSLKVTGEA 332
    ||| : : : : : ||| : : : : : ||| : : : : :
DQ 202 ----DLNVLVSLNLSNRTLVNLGVEDLVNLQELNVSAN--KALEDISQVAPVLVKEI 255
    ||| : : : : : ||| : : : : : ||| : : : : :
QY 333 SFDNDLSQLMFDGLEKIESEAFTPGPDHYNRVLVMTKSKNPSPGLATENTVYND 392
    ||| : : : : : ||| : : : : : ||| : : : : :
DQ 256 SAGCCKIKTLEL-----KNPAGAVL----- 275
    ||| : : : : : ||| : : : : : ||| : : : : :
QY 393 KSLWQESPETDYKWLBEEDTYQKNSVTGSPNKGLOKVKKNKNLEIP-----KQHNG 444
    ||| : : : : : ||| : : : : : ||| : : : : :
DQ 276 -----PEL-----ETFYQENDITNLITS--LAKLPKLNLYIKGNASLKSLETNG 319
    ||| : : : : : ||| : : : : : ||| : : : : :

```

Qy	445	VT-----ITEIGD-----NAFRNV-DFQNKTLRKYDL 470
		: :
Db	320	ATQLQIDASNCTDLTGLD1SGLSELM1QLSGCSKLEITSLKPLNPLNVITDSCAI 379
		: :
Qy	471	EEVKLPSTIRKIGAPAFQSN-NLKSPEASDDLEEIK-----EGAFMNN-----RIETL 517
		: :
Db	380	EDGCTLNNLPKQTLVLSDNENLTNITATDLPQLKTLTLDGCGTISIGTLNLPKLEKL 439
		: :
Qy	518	EKKDKLVITIGDAAFHINHHIYAIVLPESVOEIGRSAPFRQNGANNLIPMG--SKVKTLIGEMA 575
		: :
Db	440	DLKE-----NQITSI---SEITDLPRLSYLDVSVNNLTITIGDLKKLPLLEWLN 484
		: :
Qy	576	FLSNRLEHLDLSEQKQLTETIPQAPF---SDNALKEVLLPASLKTITREBAFKQHLKQLEV 632
		: :
Db	485	VSSNR-----LSDVSTLTTFPPLSYLNININNVIRTVGKMTLPSPLEKFEYAQNNSISDISM 539
		: :
Qy	633	ASALSHI-----AFNALDDNDGDEQPDNKVVVTHNSYALADGEHFIVDPDKLSSTIV-- 686
		: :
Db	540	IHDMPNLRKVDASNNLTITWIG--TFDN-----LPKQLSLDVHSNRTITSTVIH 585
		: :
Qy	687	DLEKILKILIEGLDYSTRQTQTQFRDMTTAGKALLSKNSLRQGEKQKFLQEAQPPGLGRV 746
		: :
Db	586	DUPSL-----ETFNAQTLNITWIG--TMDNLPD-----LTVNLSFNRI 622
		: :
Qy	747	-----DLDKAJAKAEKALVTYKATNGOLLERSINKAVLAYNNSAIIKANVYKRL 796
		: :
Db	623	PSLAPIGDLPLNETLTLIVSDNNSVYLSRLGTMDG-----VPKLRILDIQ 664
		: :
Qy	797	KELDLTLGLVEGKGLPAQATWQGVVLLKTPLPPEYVIGLVNVPDK-SG-----KLIYA 850
		: :
Db	665	NNYLVNTYGTGEGNLSLSDLT-----NLTELNRNNVYIDDISGLSTLSRLIY- 711
		: :
Qy	851	LDMSDTIGEGQXDAGNPTLVNDEDN-EGYHALA-----VATLADYEG 892
		: :
Db	712	-----LNLSNKIEDISALSNTLNQLBUTLENNKNTENISALSDLEN 752
		: :
Qy	893	LDIKTILNSKLSQLTSIROVPTAAYHRAGIFOA-----IQNAABAEQULLPKP 940
		: :
Db	753	LNLKLVVSKNKIIDISPVANN---VNRGAIVTASNQTYTLPTVLVSQSSFTIDNPVIWD 808
		: :
Qy	941	GTHSEKSSSESSEANSKD 957
		: :
Db	809	GTLAPSSISGNSGNYKD 825
		: :
RESULT 27		
T18353		
protein p97 - Mycoplasma hyopneumoniae		
C;Species: Mycoplasma hyopneumoniae		
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004		
C;Accession: T18353; T03805		
R;Hau, T.; Artiushin, S.; Minion, F.C.		
J. Bacteriol. 179, 1317-1323, 1997		
A;Title: Cloning and functional analysis of the P97 swine ciliun adhesin gene of Mycopla		
A;Reference number: Z18886; MUID:97175562; PMID:9023217		
A;Accession: T18353		
A;Status: preliminary; translated from GB/EMBL/DBJ		
A;Molecule type: DNA		
A;Residues: 1-1108 <HSU>		
A;Cross-references: UNIPROT:Q49542; EMBL:U50901; NID:g1399524; PID:g1399525; PIDN:AAB478		
R;Hau, T.; Minion, F.C.		
Gene 214, 13-23, 1998		
A;Title: Molecular analysis of the P97 ciliun adhesin operon of Mycoplasma hyopneumoniae		
A;Reference number: Z15097; MUID:98391007; PMID:9729120		
A;Accession: T03805		
A;Status: preliminary; translated from GB/EMBL/DBJ		
A;Molecule type: DNA		
A;Residues: 1089-1108 <HS2>		
A;Cross-references: EMBL:AF012905; NID:g2654768; PIDN:AAC32526.1; PID:g2654769		
A;Experimental source: strain 232A		
C;Genetics:		
A;Genetic code: SGCS		
A;Note: p97		

Db 455 IMGHVDHGKTSLDKIRKRVVTEAGGITQIHGAYWVEKNKWSFIDTPGHEAFSOMR 514
QY 555 QNGAN---NLIFM---GSKVKTLMGEM-----AFLSNRLHLDLSEKQQLTE-- 594
Db 515 NRGQVTDIAVIAADGGVKQOTIEALBHAKAANVPVIFAMNKMMDKPNVNDPKLKAECA 574
QY 595 -----IPVQAFS-----DNALKEVLLPASLKTIREEAFKKNHLKQLEVASA 635
Db 575 ELGYNPVDWGGHEHPIVPSAKTGGIDNLETLIQADIM-----ELKAIEGSA 624
QY 636 LSHIAFNALDDNDGBOFNKVVVTHNSYALADGEHFIVD-----PDKLSSTIV 686
Db 625 RAWLVGSGVEKGRG-----AVATVIVQSGTSLVSGDSFFAETAFGKVRVTMTDDCKSIQ 677
QY 687 DLE-KILKILIEGLDSTLRQTQTQPRDMTWTAGKALLSKSN-----LRQGEKQKLOEQA 740
Db 678 NUKPSVALITGLS-----EVPFAGSVLIGVENDSIARLQAKKATYLRO-- 722
QY 741 PFLGRVLDLKAIAKAEKA-----LVTKKATKN-----GOLLERSINKAVLAYN 783
Db 723 -----KALSKSTKVSFDELSEMVANKELKNIPVIAIKADTQGS--EAIKNSILLEIN 771
QY 784 NSAIKKANVNR-----LEKELDIL-----TGLVBGKGPLAQATWQGVYL 823
Db 772 NEEVAIQVTHSGVGGITENDLSVSSSEHAVILGFNIRPTGNVONKA----- 818
QY 824 LKTPPLPPEYIGLVNVPFDSKGLIYALDMSDTIGEGQD---AYGNPILNVD-----ED 875
Db 819 -----KEYNVSITKY-----TVIYALI-----EGMRSLLLGLMSPIIEEHTGQAEV 860
QY 876 NEGYHALAVATLA 888
Db 861 RETFNIPKVGTTA 873

RESULT 29
T30822
Impl protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30822
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infect. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis PG21 deletion mutants by cultivation in the presence of a specific inhibitor
A:Reference number: Z19884; MUID:95369882; PMID:7543881
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: UNIPROT:Q49525; EMBL:U21962; NID:G790243; PID:G790244; PIDN:AAA81013
C:Genetics:
A:Gene: Impl
A:Genetic code: SGC3

Query Match 3.8%; Score 191; DB 2; Length 1365;
Best Local Similarity 20.0%; Pred. No. 0.47;
Matches 185; Conservative 154; Mismatches 315; Indels 270; Gaps 43;

QY 15 VSVVTHNQEVSLVKEP-----ILKQTO-----ASSISGADYAE-----SSGKSKLKIN 59
Db 258 VABITKLETFNKDEAKFNELKQTRNQIQEFINTKNPNYSSELISQTSKRDSSKNSVT 317
QY 60 ETG--GPVDDTDTDLFSD-KRTTPEKIK-DNLAKGPREQELKAYTE-NTESKQITSGSQ 114
Db 318 DSSNKSIDIESANTELKQALAKANADKVQADNLAKSIKEQLNNSVSNANTLSAKLTDKONT 377
QY 115 LQSKESLSLNTKVPSTSNWEICDFTTKGNTLVGLSKSGVEKLSTQDHLVLPQADGTO 174
Db 378 IQAK--TELEKV-----QKADQAIKSNNTASM-----QSAKSSLDARVAEITK 420
QY 175 LIQVASFAPTPDKTAIAEYTSRAGENGESIQLDVGKEIINEGEVFNYS--LLKKVTIP 232

421 KLE-----TFNKDKAEKF-----NELKQTRNQIQEFINTKNPNYSSELISQLT-- 464
QY 233 TGYKHIGODAFVDNKNIAEVLNLPESLETISDYAPFAHLALKQIDLDPNLKAIGE----- 285
Db 465 --SKRDSKNSTVSDSNKSDI---ESANTELKQALAKANADKVQADNLAKSIKEQLNNSVS 519
QY 286 -----LAPFDNOI--TGKLSLPRQLMRLAERAFKSNHIKTIEPRGNSL--KVI----- 329
Db 520 NANTLSAKLTDKONTIQAKTELEKEVQK-ADQAIKSNNTASMOSAKSSLDARVAEITK 578
QY 330 -----GEASFQDNDLSQLMLPDGLEKIESEAFTCGNPGDDHNNRVVLTWTKSGKNPSGL 382
Db 579 LETFNKDKAEKF--NELKQT-----RNQIQEFINTKN----- 609
QY 383 ATENTYVNPDKS--LWQESPEIDYTKWLEEDFTYQKNSVTGFSNK-----GLQ 428
Db 610 -----NPNYSSELISQTSKRD-----SKNSVTSSNKSIDIESANTELKQALA 651
QY 429 KVKRNKNLEIPKQHNGVTTITEIGNAFRNVDFQNKTLRYD--LBEVK--LPSTIRKIGA 484
Db 652 KANADK---VQADNLAKSIKEQLNNSVSNANTLSAKLTDKONTIQAKTELEKEIQKANQ 708
QY 485 FAFQSNLKSFEASDDLEIEKEGAPMNRRIETLELKDVLVIGDAAFINHIYIV----- 540
Db 709 -AIKSNNTASMOSAKSSLDARV-AEITKLETFN-KQKEAFNKLKQTRNQIQEFINTKN 765
QY 541 -LPESVQEIGRSAPRONGANNLI----- 562
Db 766 NPNYSSELISQTSKRDSSNVTSSNKSIDIESANTELKQALNTAKAKKSSIDNLRPLK 825
QY 563 -FMGSKVKTG-----ENAFLSNRLHLDLSEKQQLTEIPVQAFSDNALKEVLLPAS--- 613
Db 826 NDLQSKIEEFGPIRNTNFSWISSKLETTKKNLAKBELTKADAIKNNPSSKQALDKSSQV 885
QY 614 -----LKTREAFKKNHLKQLEVASALSHIAFNALDDNDGBOFNKVVVTHNSY 666
Db 886 QKLGNELLKITEE--FGKVETKNSNIGYRLFKLA-----QAEQFNNS----- 926
QY 667 ALADGEHPIVDPDKLSSTIVLEKILKILIEG-----DYSTLRQTTQTPRDMTWTAGKA 720
Db 927 -----DVKLKNABEKSQTLSSKQKLGQNSTKDYLT-QLSTEMSTQESTIKKVI 975
QY 721 LLSKSNLQGEKQKPLQBAQFPLGRVDLDLKAIAEAKALVTKKATKNQQLLERSINKAVL 780
Db 976 VNIQAHIRNNLSQY-----RLEADKLIANKRGYGVGDKVGIESLKQWDLMDSDSVL 1026
QY 781 AYNNSAIKKANVKRLEKELDLLTG 804
Db 1027 SVDDSLKDDFN-----KALRVLVG 1045

RESULT 30
T50073
myosin-like coiled-coil protein sp8 [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C:Accession: T50073
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.
submitted to the EMBL Data Library, December 1999
A:Reference number: Z25034
A:Accession: T50073
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1727 <MCD>
A:Cross-references: UNIPROT:Q9UTK5; EMBL:AL133357; PIDN:CAB62414.1; GSPDB:GN00066; SPDB:
A:Experimental source: strain 972h(-); cosmid c1486
C:Genetics:
A:Gene: SPDB:SPAC1486.04C
A:Map position: 1

Query Match 3.8%; Score 190.5; DB 2; Length 1727;
Best Local Similarity 19.7%; Pred. No. 0.7;
Matches 215; Conservative 177; Mismatches 406; Indels 293; Gaps 49;

Db 1503 F---GDIASVKAQYQVQVNLADSSGNOAIPOITITINVD---TVGPVIAKADNVSHVNT 1556
Qy 542 PESVQIGHSAPRONGANNILPMGSKVKTLGEMAFSLNRLEHLDSL---OKQTEIPVOA 599
Db 1557 TKTEAEFFODARLDVTDNN-----DDTDLITISFAEKVNLNPKQKYEYITATDT 1608
Qy 600 FSDNALKEVLLPASL-KTIREBAFKKHLKQLEVASA---LSHIAFNALDNDGDEQDNK 656
Db 1609 KGNQTTKEITVQVSKDKPVITADPKISYQKIEVTEANFLSGVHAETVDELGDVKTISD 1668
Qy 657 VVVKTHN---SYALA-----DGEHFIVDPDKSSTIVDLKILKLEGLDYSLRQTQT 709
Db 1669 FAEKVDENKVGTVTVTLNAKDEYGNTPAEVPKVSUSI--FNKIAPTNNADNKTIEAVNEL 1726
Qy 710 QFRD---MTTAGKALLSKSNLROGEKQFLQBAQFFLGRVLDLKAIAKAEKALVTKATK- 766
Db 1727 PSLSEIFKTEAKDYLGSNKL-----KVITYPE-QTIKGNVPGBYSI-----KVTTKD 1772
Qy 767 -NGQLLERSINKAVLAYNNSAIKKA-NVKRLE---KELDLL-----TGLVEG---KGP 811
Db 1773 DSGNIAETVTLTIKDTGSPISKWTSTKLEVSQKEPNWITFFGKATDIWDGVTKNI 1832
Qy 812 LAQATMVQ---GVYLLKTPPLPEYYIIGLVYFDKSGKLIYALDMSDTIGEGOKDAYGN 867
Db 1833 KVDSSSEVNLKVGTY-----PIYFTVTDALGNESSKLSKSTVQIVDT-----SS 1875
Qy 868 PILVDEDEGHALAVATLADYGL-DIKT-ILNS--KLSQTSIRQVPTAAYHRAGIF 923
Db 1876 PELTIDKLBISYP--TGKTIQSKQLQDITGKVTNSYGTVKVTTNLSKI--VDWNAKAGY 1931
Qy 924 OAIONA-----AAEAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKNRRGHSAILPR 977
Db 1932 KVTVTATNSSGGVAEKTIILL-----TVKNSDSSFIAPVPSKD-----DNKNPKAKNI 1979
Qy 978 TG 979
Db 1980 TG 1981

RESULT 35
T18354
adhesin - Mycoplasma hyopneumoniae
C:Species: Mycoplasma hyopneumoniae
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18354
R:Wilton, J.B.; Scarman, A.L.; Walker, M.J.; Djordjevic, S.P.
Microbiology 144, 1931-1943, 1998
A:Title: Reiterated repeat region variability in the ciliary adhesin gene of Mycoplasma
A:Reference number: 218887, MUID:98361039, PMID:9695926
A:Accession: T18354
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1092 <WIL>
A:Cross-references: UNIPROT:O07132; EMBL:AF001398; NID:g2150123; PID:g2150124; PIDN:AAC3
C:Genetics:
A:Genetic code: SGC3

Query Match 3.7%; Score 186.5; DB 2; Length 1092;
Best Local Similarity 18.7%; Pred. No. 0.54;
Matches 203; Conservative 159; Mismatches 368; Indels 353; Gaps 48;

Qy 20 HNGEVSFLVKEPILKQTOA---SSSISGADYAESSGSKLKNETSGPVDVTDVLFSDK 76
Db 142 HNGDI---AKSDIYEQTAVAFKQSNLLVAEFNPSLKKITEKLNQQLNLSKTIKN-FADE 197
Qy 77 RTTPEKIKONLAKGPREQELKAVTENTESEKQITSSQLEQSKESLSLKNTPVSTSNWEI 136
Db 198 KTSQQDPSTLRADIFQYDLN-TARNAE----- 224
Qy 137 CDPITKNTLVGLSKGVKLSQTDHLVLPQSAADGTQLIQVAFPTPKKTAIAYETS 196
Db 225 -DLIDIKLANYFPVLKNIINLNNAENKLPN-----NLGNIFEFSAKDSST--NQVYS 275

Qy 197 RAGENGBI---SOLDVDCKEIINEGE---VFNSYLLKKVVTIPTGYKHIGODAFVDN--- 246
Db 276 IQNQPISFLKADILSOSAREILASDPVQPIN--ILRLMKDNSSYFLNPFEDFVNNLT 333
Qy 247 KNIAEVNLPESLETISDYAF-----AHLALKQIDLPDNLKAIGELAFPDN 291
Db 334 KMQKEDLNAGQNLSAVEFLADIKSGFFPGDKRSSHTKABEISNLLNKKENIYDFGKYNG 393
Qy 292 QITGKLSIPROLMRLAERAFKSNHKTIEFRGNSLKVIGEASFQDNDLSQLMLPGCLEKI 351
Db 394 KFNDRLNSP-----NLEY---SLDA-ASASLDDKDKSIILIPYRLE-I 431
Qy 352 ESEAFPTGNPGDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSL-----WQSPSEID--- 403
Db 432 KDKFF---ADLLYPD-----TKDNILVKEGILKLTGFKGPKLDLNI 471
Qy 404 -----YTKWLEEDFTYQKNVSTGFSNKGLOKV-----KRNNKLEIPKOHNGV 445
Db 472 NQOIPKTEVLPFFKEGKEEQAKLDYGNILNPYNTQAKVEVEALFGKNKQOEIYQALDGN 531
Qy 446 TITEIGDNARNV-----DFQNTLRYDLEEVKL-----PSTIRKIG 483
Db 532 YAYEFG--APKSVLNSWTGKIQHPEKADIQRFTRH---LEQYKLGNSVLNQPOPTKEQV 586
Qy 484 AFAFOSNNLKSFASDDLEIEKEGAFMNRITLLELKDVLVTIGDAAFHINHIYAI--- 539
Db 587 ISSLSKNFP-----FNGHQVASYQDILLTKDKL-TVLETLVDLAKKWLGNR 634
Qy 540 -VLPEVQIEIGRSAPFRQNGANNLIPMGSK-----VKTGEMAFSLNRLEHLDISEQKQ 591
Db 635 AQFPKEVQYTKDIFAE--ADKLKFELEGKKDPYVQIKIEIHQLSP--NILARNDIK--- 687
Qy 592 LTPVQAFPSNALKEVLLPASLKTIRBEAFKKNHLKQLEVASALASHIAFNALDNDGDE 651
Db 688 -----SDGYGVLLLPQSVKT---ELEGKNEAQ-----IFEAL----- 717
Qy 652 QFDNKVVVTHHNSVALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSLRQTQTQTF 711
Db 718 -----KYSLEIENSAF-----KTTIILD-----KNLEGTDFKT-----F 746
Qy 712 RDMTTAGKALLSKSNLROGEKQFLQEAQF--FLGRVLDLKAIAKAEKALVTKATKNGQ 769
Db 747 GDFL-----KAFLLKAAQFNFPAPWAKLDDNLIQYSFEAIKKGTTKEG- 789
Qy 770 LLEERSINKAVLAYNNSAIKKNVRRLEKELD-LLTGLV-----EKGPLAQATMVQ 820
Db 790 -----KREBVDKVKELDNKIKGILPQPPAPKPEAAKPVAAKPEAAK 831
Qy 821 VYLLKTPPLPEYYIIGLVYFDKSGKLIYALDMSDTIGEGOKDAYGNPILNVDENEG-- 878
Db 832 PETTKPVAAPKE-----AAKPVAA-----KPVAAKPVATNTNTNTGFS 869
Qy 879 -----YHALAVATLADY---EGLDIKT-----ILNSKLSQTSIRQVPTAAVH 918
Db 870 LTNKPEKEDYFPMAFSYKLEYTDENKLSKTPEINVLVHQSEYEEQKILKELDKTVLN 929
Qy 919 RAGIQAIONAAAEAOQLLPKPGTHSEKSSSESANSKDRGLQSNPKNRRGHSAILPR 978
Db 930 LQYQOEKVTSEYQKLSHPMTTEGSPNOGKKAEGAPNOGKKAEGAPSOQKKAEGAPNQ 989
Qy 979 GSK 981
Db 990 GKK 992

RESULT 36
BWEYDL
RAD50 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein N0872; protein YNL250W
C:Species: Saccharomyces cerevisiae
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Jul-2004
C:Accession: S05808; 563223

280	ALKDQAA TKQ-----AQNTLSSKEELKGHGKGLN-----PANFTP-DYYKLTUEQEQAM	330
352	ESBA-----FTGNPGDDHYNRVLVLTGSGNPSGLATENTYVNPDKSLMQESPEIDYT	405
331	EKBALALNKKVFPENQADAAKVTEMI-----NVKNP-----TEK-----QKQMSDV	372
406	KWUEEDFYQKNSVTGFSNKGLOKVK-RKNKLEIPKQHGNVITITIGDNAPRNVDFONKT	464
373	VGLJNDVR-----EKLGLQKUKISNQAKF-----AWDVAJDNPK	408
465	LRKYDLBEVLKSTIRKICAFQSNLNKLSFEASDDLEIEKEGAFMNNRIETLELKDCLV	524
409	EFDHVNAINR-----AAKENGFKFPGQNFYENLSWG-----RFTTQEGKVSUY	453
525	TIGDAAFHINHIYAIVPESVOEIGRSAPRQ-----NGANLLIFMGSKVITLGE MAPL	577
454	DFHKAARN-----ALVSMMLNDCHSGYSHLDSLLDANETN-----MAVSI SGLNDI	500
578	SNRLEHLDSBKOLTEIPVQAFSNALKEVLLPASLKTIREAFKKNHLKOLEVASALS	637
501	SAKI-HIISYNQTKLVEA-----NTVEEGTAP-----VPKSKETLOKEVATNOE	543
638	HIAFNALDNDGDGE-QFDNKVVVKTTHNSYALADGEHIFVDPKLSSTITVDLEKILKITE	696
544	KLATAQQAESDAQARSQAALNTAKTTQATAEKELSV-----HKATLANIJEV-----	593
697	GLDYSTLRQTTQTFPRDMTTAGKALLSKNSLBOGEKQKFLQBAQFPLGRVLDLDAKAE	756
594	-----ATKSTTNYBEKVRQTAA-----AETNLQOTKDO-----LATTNELI--QN	631
757	KALVTKKA-TK--NGOLLERSINKAVLAYNNSAIK--KANVKRLSEKELDLTLGLVEGKGP	811
632	RAVVLKRAKTVADAQAEIQOTSAK-VLKEQBAQKAEENTNSLKEVLDLAKENLNQK-Q	689
812	LAQATWQGVVLLKTPLEPYIIGLVYVFDKSGKLIYALDMSDITIGEGQKDAGNPIILN	871
690	VALKSTSTRSLARLENAQPTYE-----KAVNELNKAAAVVQAKEAYENLSKS	736
872	VDENEGVHALAVATLADYEGLDIKTILNSKLSQLTSTRQVPTAAYHRAGIQATQNAAA	931
737	LEBLKE---QOAVATLAVTQAOEDLSNAKLELQYQGLTRDLBQAQAEQRQENLQEQVA	793
932	EABQLLPKPGTHSEKSSSESANSKDRGLQ-----SNPKTNRGHS-----	972
794	KEQORLEERACRQMWLASATSADKTPGLQQLSFSKQKEQPAQOALTHTHSEPKTKQVAKA	853
973	-AIIIPRTGSKGSFVGIILG-YTSVALLS	998
854	QESLPHTGEQKSIWLTIFGLFMVAGAIS	881

RESULT 38
G69708

Chromosome segregation SMC protein - *Bacillus subtilis*
N/Alternate names: minichromosome stabilizing protein SMC
C/Species: *Bacillus subtilis*
C/Date: 05-Dec-1997 #sequence revision 05-Dec-1997 #text_change 09-Jul-2004
C/Accession: G69708 J04819; F04029
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier
iech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holasappel, S.; Hosono, S.; Hulio, M.P.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinio,
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mause
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rome, M.; Sadaie, Y.; Saco, T.; Scanlon,
A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpest, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A/Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.


```
Db 610 -----ELKAIEGSAVAVLESGVGRG-----AVATVIVGSGTSLVSGDSPAETAP 657
Qy 678 -----PKLSTIIVDLKILKLEGLDYSTLRTOTQTRDMTATAGALLSKSN----- 726
Db 658 GKVRTMTDDQKSIQNLKPSMVALITGLS-----EVPAGSVLIGVENDSIA 704
Qy 727 -LRQGEKQKFLQBAQFFLRVDLDKAIKAEKA-----LVTKKATON-----G 768
Db 705 RLQAKRATYLRQ-----KALSKSTKVSFDELSEMVANKELKNI PVVIKADTQG 753
Qy 769 QLLERSINKAVLAYNNSAIKKANVR-----LEKELDLL-----TGLVEG 808
Db 754 SL--PAIKNSLLLENNBEVAIQVHSVGGITENDLSVSSSEHVAIVLGFNPRTGNVN 811
Qy 809 KGPLAQATWQGVYLLKTLPLPEYIYGLNVYFDKSGKLIIVALDMSDTIGBGQKDAYG-- 866
Db 812 KA-----KEYNVSIKTY-----TVIVAL-----IEEMRSELLGLM 841
Qy 867 NPILNVD-----EDNEGYHALAVATLA 888
Db 842 SPIIEEHTGQAEVRETFNIPKVGTTA 868

RESULT 42
P96673
Hypothetical protein F13011.30 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: P96673
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizlar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: P96673
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1313 <STO>
A:Cross-references: UNIPROT:Q9XIP6; GB:AB005173; NID:G5042434; PIDN:AAD38273.1; GSPDB:GN
C:Genetics:
A:Gene: F13011.30
A:Map position: 1

Query Match 3.6%; Score 182.5; DB 2; Length 1313;
Best Local Similarity 20.2%; Pred. No. 1.1;
Matches 211; Conservative 171; Mismatches 380; Indels 281; Gaps 51;

Qy 31 PILKQTQASSISGADYAESGSKLKINETSGPVDVTDLPFSKRTTPEKIKDNLAKG 90
Db 22 PRLSKLSASKDS-----NSAPVHSLRVKG-----TELQTLNQIQEDLKA 65
Qy 91 PREQEL-----KAVTENTESEKQITSGSQLESLSKNTVPSTGNWEICDFTIKGN 144
Db 66 DEQIELKKDKAKAIDDLKSEKLVBEAN--EKLEALAAQR--AESPEVEKF----- 116
Qy 145 TLVGLSKSGVEKLSOTD-----HLVLPQADGTQLIQVA-----SFAFTPDKK- 188
Db 117 RAVELEQAGLEAVQKDVTSKNELESIRSQHALDISALLSTTEELQVRKHLSMTADAKN 176
Qy 189 -----TATAEYTSR-----ACENGEISOL--DVDGKEIINEGEVFNLSLLKVTIP 232
Db 177 KALSHAEATKTAETIAHKAETIAGELRLKALLGSKBEKEAI--EGNEIVSKLKSIEILL 235
Qy 233 TGYKHIGQDAFVNDKNIAEAVNLPESLETISDYAFALHAKQIDLPDNLKAIGELAFPNQ 292
Db 236 RGE-----LEKVSILESSLKE-----QBLVEQLKVDLEAA----- 266
```

```
Qy 293 ITGKLSLPRQLMLAEAFKSNHIKTIEFRGNSLKVIGEASPDNDLSQLMLPDGLEKIE 352
Db 267 -----KMAESCTNS-----SVEEMKNKVHEL--EKEVESNRSKSSASESMESVM 309
Qy 353 SEAFGNPGDDHYNNRVVLTGSKGNPSGLATENTYVNPDKSLWQESPEIDYTKWLBEDF 412
Db 310 KQLAE-----LNHVLEHTGS-----DNAAQKEKIEL-----LEXTI 340
Qy 413 TYQNSVTGFSNK--GLOKVRENK-----NLEIPKQHGVTIITEIGN--AFRNV 458
Db 341 EAQRTDLEEGROVCIKAEKASLENLVESIKSLEISQEBK-----TRALDNEKAATSN 396
Qy 459 D--FQNTLRKYDLEVKLPSTIRKIG-----APAFQNNLKSFPASDDL-----REIKEGA 508
Db 397 QNLDDORTELSIELERCKVEBEKSKDMESITLALQEAESTESSEAKATLLVCBELKN-- 454
Qy 509 FMNRIETLELKDKLVT-----IGDAFPHNHIVAVILPESVO---EIGRSAPRQNGAN 559
Db 455 -CESQVDSLKLASKETNEKYLEMLEDARNEIDSLKSTV--DSIQNEFENSKAGHEQ---K 508
Qy 560 NLIFMGSKVKTGEMAFLSNRLEHLDSLSEKQKLTETIPVQAFSDNALKEVLLPASLKTI 619
Db 509 ELHLMGCVKGS-----EEENSSQBEVSRV-----NLLKESEBDACARK--EE 550
Qy 620 EAPKKNHLKQLEVASALSHTAFNALDDNDGDEQFDNKVVVKTTHNSYALADGEHFIVDP 679
Db 551 EASLKNLKVAB-----GS-----VKYLOETLGEAKAEM----- 580
Qy 680 KLSSTIVDLKILKLEGLDYSTLRQ---TTOTFRDMTATAGALLSKSNLRQGEKQKFL 736
Db 581 KLESLLDKBEDLNVT--ELISSUREWEGSVLEKIEBELSKVESLVK---ETKLQSIIT 635
Qy 737 QEAQFFILGRDLDKAIKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKKANVRKLE 796
Db 636 QEAEELKGR-----EAAHMKQIEELSTANASLVDEATKLQISVQESDELKEKEAGYLK 688
Qy 797 K--EIDL-----LTGLVSGKPLAQATWQGVYLLKTLPLPEYIYGLNVYFDK 843
Db 689 KIBELSVANESLADNVTDLQSIQVQESKDLKEREV---AYLKK-----IEELSVANESLV 741
Qy 844 SGKLIYALDMSDTIGEQKDAYGNPILNVDSNDSGYHALAVATLADYEGL--DIKTILNS 901
Db 742 ETKLQH-IDQAEELRGREASHLKKIEBELSKENEN-----LVDNVANMQNIAEESKDLRER 796
Qy 902 KLSQTSIRQVPTAAYHRAGIFQAIQNAABAEQLPKPGT---HSEKSSSSSESANSKD 957
Db 797 EVAYLKKIDELSTANGTLADNVNVLQNISEENKELRERETTLKKABELSELNLSLVDK 856
Qy 958 RGLQSNPKTN---RGRHSAILPR 977
Db 857 SKLQTVVQENBELRERETAYLKK 879
```

RESULT 43

```
T43527
sp8 protein - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Feb-2000
C:Accession: T43527
R:Jimenez, M.; Goday, C.
submitted to the EMBL Data Library, June 1997
A:Description: Schizosaccharomyces pombe myosin-like.
A:Reference number: Z22545
A:Accession: T43527
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1173 <JIM>
A:Cross-references: EMBL:AF010473; PIDN:AAB65416.1
A:Experimental source: strain 972h(-)
C:Genetics:
A:Gene: sp8
```

```
Query Match      3.6%; Score 182; DB 2; Length 1173;
Best Local Similarity 20.2%; Pred. No. 0.98;
Matches 199; Conservative 163; Mismatches 326; Indels 296; Gaps 48;

QY 33 LKQTOASSISGADYA-----ESSGSKLKINETSQVGD-----66
Db 281 LHSLOKSEKSGEYSRIKTLMLKQSESDNRKLLDNQOMWEIKQLBNGVIELEKOR 340
QY 67 -DTVDFSDKRTTPBEKIDNIAKGPREQELKAVTENTESE-----KQITSSQSLR- 116
Db 341 FSTLEAKFTQOKNTSYSEREALLESSL-SDLOKSHTSLESQYNSLRNIEQLOQAASKLAE 399
QY 117 -----QSKSLSLKNKTVTPSTSNWEI-----CDFITK-GNTLV 147
Db 400 EMVERVKTEYDEYRLQTSSELSKHNKIKITSLEQRIIVILQDEIATSSSLRCNITKQSETRV 459
QY 148 GL-----SKSGVEKLSQTOH---LVLPQNAQDT-----OLIQVAS 180
Db 460 ALLLEENKHNLSHRNAEKQHLEK--ENDYKQQLLVTEDLRKTREDEYKELLRHAD 517
QY 181 FAPTDPKTAIAEYTSRAGENGESIQDVGKEIINEGEVF-----NSYLLKKV 229
Db 518 ARSTLQK---LREDYTKALEQVEDINKEIATKAGINESQFPPISEKEDPLRQEVYVLK- 573
QY 230 TIPTGYKHIGQDAFV-----DNKNIAEYNLPE-SLETISDYAFALHAKQIDLPDLNKA 282
Db 574 -----QNAMLLTQLQSSNLNFAEITSPSPDLDSV-----MKLGSLDLQNHVKR 616
QY 283 IGLAFPDNQITGKLSLPQLMRLAERAPKSNHIKTIIEPRGNSLKVIGASQDNDLSOL 342
Db 617 ISK---EMEI---ISCORQLLFLENKULK---RTVE---SSNRVIAD-----651
QY 343 MLPDGLKEIESAFTGCPGD-DHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLMWQESPE 401
Db 652 -JORGTEKDVSTSESVGERSYLNWVALLNESNKS-----LRENLERN-EEVITELREK 705
QY 402 IDYTKWLEEDFTYQKNSVTG--PSNKGLOKVRKNKLEIPKOHNGVTITEIGDNAPRNV 459
Db 706 IETLKTDLANFRLNKEQLSSQLQTEKAAVKKLENSNEEY-KRHNQEIILLSNSSTSSD 764
QY 460 ----FQNKTRKYDL-BEVKLPSTIRKIGAPAFQSNMLKSEASDDLEETKEGAFMNRNIE 515
Db 765 ASRLKNELVSKENLIBELN-----QBIGHLKSELETVKS--KSEDLLE--NERAQNSKIE 815
QY 516 TLRLKDKLVITIGDAAHINHIAIPLPESVQETGRSAFRQNGANNLI FMGSKVKTLGEWA 575
Db 816 QLELKNTKLA---AAWRTK--YEQVVKNSLEK--HNQIRQQLSQKTSSELEAKVAECHOJN 868
QY 576 FLNSR-----LEHLD-----LSEKQKLTET-----PVQAFSDN 603
Db 869 EQLNKPSATPTATQSEPTSVLSLEEFNSTKEELSSSTQKLSLEIMDILNTTKEELEKVRQN 928
QY 604 ALKEVLLPASLKTIRREAPKKNHLKQLEVASALSHIAFNALDDNDGDEQFKNKVVVKTH 663
Db 929 SNKSEKTSKDEIPIPEEERKMKVMOQEVRLRSRIA-----KELQKNELLARK---976
QY 664 NSVALADG---EHFIVDPDKLSSTIV-----DLEKILKLIBEL-----DYS 701
Db 977 QNOVLQDQVKALQETVVSSEEAESASVHADTKDLENLKKTEEMLSVTFQVIFNESISDFS 1036
QY 702 TL-----RQTQTOFRDMTTAGKALLSKSLRQEKQKFLQ 737
Db 1037 TSTADPTFTTQKEWEKREILQKQVEEQASHQKQOLDNIRKELENRNKLKLSMLEKNLA 1096
QY 738 EQAOFFLGRVDLKAIAKASKALVTKATKNGOLLERSINKAVLYNNNSAIKKANVKRLEK 797
Db 1097 RV-----RALEQSKKKDPAIILSLEASKN-----TDSNK-----SNSEVPAQAQVK--EK 1139
QY 798 ELDLLTGLVEGKPLAQATWQGV 821
Db 1140 KLIATKTHSVDTNSPPKRSSSDAGM 1163
```

```
RESULT 44
AE1717
C:Species: Listeria innocua
C:Strain: Listeria innocua (strain Clip11)
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AE1717
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Bloeker,
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournan, A.; M
ok, C.; Schlueter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669.
A:Accession: AE1717
A:Status: preliminary
A:Molecule type: DNA
A:Accession: 1-1622 <GLA>
A:Cross-references: UNIPROT:Q929J4; GB:AL592022; PIDN:CAC97509.1; PID:gl6141793; GSPDB:
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin2281

Query Match      3.6%; Score 182; DB 2; Length 1622;
Best Local Similarity 18.9%; Pred. No. 1.6;
Matches 240; Conservative 165; Mismatches 436; Indels 430; Gaps 58;

QY 2 KKHKTVALTLTVSVVTHNQEVFSLVK-EPILKQTOASSISGADYAESGSKSKLINE 60
Db 64 KKTGETKEVSTEDGTIVEN---SLSEGEYIVKEKKAAPGT-----LDEQTYNV 110
QY 61 TSGPVDVDTWDLFSDKRTTPEKIDNIAKGP-----REQELKAV-TENTESQKITSG 112
Db 111 TLADKDEVITSI-----STKDEAKTPIVTEQPSKKNLKAIVADIADNIFAVKENG 161
QY 113 SOLEQSKES-----LSLN-----KTV-----PST- 131
Db 162 TGNELSTTNRINKGAVVLKNNFTFSGKNYKAGDTFKVLPTFNFGTTLNLDGFLPSTE 221
QY 132 SNWE-----ICDFTKNTVLGLSKSGVEKLSQTD-----H 162
Db 222 AKWDLNASTRELITITFLKDGVOEGNYDIELS-TALKSFSETDKTRQEAVENTAGNTIYQ 280
QY 163 LVLPQAAADGTO-LIQVASFAFTPKKTAIAEY--TSRAGENGESIQDVL--DGKEINE 217
Db 281 LEIIEVDKATQVLMLEATPGKLPKATVDARFNTKESNELGELRLSDTAYGSTIINR 340
QY 218 G--EVFNSYLLKCKVTIPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFALHAKQID 275
Db 341 NSIKVYSTDISAKGT-----FIGSKQL---LNEGTDYELIYAPSGLTIK---381
QY 276 LPDNLKATGELAFPDNQI-----TGKLSLPQLMRLAERAPKSNHIKTI-----FRGNSL 326
Db 382 LKEGLKAKGYQVITYERSIDKNTSITTLTGSATTVSSGMSLNGSMITISVTIKAYDHLTK 441
QY 327 KVIGEASFQ-----DNDLSQMLPDGLE-----KIESEAFGTGNP---360
Db 442 KAVNPVTCIDMTWNTNVYDLANLTPDTVLKDVLTDDNVYVADSLKIKQVTFNEESGEV 501
QY 361 --GDHYNRVVVLWTKSGKNPSG-----381
Db 502 VIGEDASND---WTVSTISDNGSFNNMYKTKKAYQVITYSTKLTDFSPRIKKEVTD 557
QY 382 ---LATENTYVNPDKSLMQESPEIDY-----TKW-----LEEDFTYQKNSV 419
Db 558 KGVKATENPDFKPD-LLKKEAGEIDYNNYNTMDTITVNSEGINMNQINIVDEFSTGVKSL 616
QY 420 TGF-----SNKGLQKVRKNKLEIPKQ--HNGVTITEIGDNAPRNVQFNKTLKPYDL 471
Db 617 ISYSVYVPSGSGYNILTEGRDFTIQKDVTPDGFKIKLIGNATTADNKIVVAKTKIDLT 676
QY 472 EVKLPSTIRKIGAPAFQSNMLKSEASDDLEIEKGAFMNRIETLELKDCLVITIGDAAF 531
```

```
Db 677 DG--ANTLNKASFYFDGSLTQYSEIKAEATPETSILAN-----GGKVGWNPATG 727
Qy 532 HNHIIYAIVLPESVQIEIGRSAPRONGANNLIPMGSKVKTLCGEMAFSLNRLEHLDLSEQKQ 591
Db 728 EINWIV-----SVNMGKK-----YKLVLI---DDBFLDGTTFVEGSLQRRNVNVSSE 772
Qy 592 LTEIPVQAFSONALKEVLLPASLTIRSEAFKKNHLKQLEVASALSHIAFNALD----- 645
Db 773 LTDLSI-----PLEIKTQAQVGDNYPTKIDTSANKHLFEGNLDNTRVVF 819
Qy 646 -----DNDGDEQ-----FDNKVVVKTTHHNSVALADG--- 671
Db 820 KYKTKPKONWFFSQWNNKAIUSDNGADEQIYETKEFAFLQNEVIKVAIGNIDSTFGNKVN 879
Qy 672 ---EHFIVDPDK-LSTIVD-----LEKILKLI-----EGLDYSTLRQ 705
Db 880 NMMLNINITPRTLSNPVITNRLEQNGTAQOFIKNSFQVINTKTNEPINEENYDITFEGN 939
Qy 706 TTQTOFRDMTTAGK-----ALLSK--SNLRQGEKOKFLQEAQFF----- 742
Db 940 FTTIQPKYNTAMAPIKNVYSTISLISGPISTNETTVEADFSNVPFFKKRNVAVSPVFTV 999
Qy 743 -----LGRVDLDKAIKAKEA---LVTKKATKNGQLLERSINKAVLAINNSA 786
Db 1000 GTGSGIATIGTIKITKVDDEDTTKLEGAKFQLYTLDEKSGQEIT-----TNS- 1048
Qy 787 IKKANVKLEKELDLTLGLVEGKGLAQATWQGVLYLKTPLPPEYVIGLNVYFDKSGK 846
Db 1049 -----EGBI-LLDGISQSKYKLVETKAPEGNY-----SDEYKEGKEITVSSSGE 1092
Qy 847 LIYALDMSDTIGEGOKDAVGNPILNVDENEGHALAVA---TLADYEGLDIKTILNSKLS 904
Db 1093 -----ELLTIKNAMKK--GNVIL--TKKDSASHEVLADAEPELQATGSKLEKLTDDAS 1144
Qy 905 QUTSIROVPTAAHYHRAGIFQAIONAAAEQLPKPGTHSEKSSSSSESANSKORGLQSNP 964
Db 1145 GNIEITDL-----APGDYKLIETKAPTGYQLDATPVNFTIDFNQSEAAKVS----- 1190
Qy 965 KTNRGHRSAIL 975
Db 1191 KNTAKTGTVM 1201

RESULT 45
AE1449
Hypothetical protein lin0132 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AE1449
R:Glaeser, P.; Frangeul, L.; Buchrieser, C.; Anand, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Sincos, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AE1077; MUID:21537279; PMID:11679669
A:Accession: AE1449
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1946 <GLA>
A:Cross-references: UNIPROT:Q92FH4; GB:AL592022; PIDN:CAC9365.1; PID:G16412551; GSPDB:C
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin0132
```

```
Query Match 3.6%; Score 182; DB 2; Length 1946;
Best Local Similarity 20.3%; Pred. No. 2.1; Mismatches 302; Gaps 55;
Matches 217; Conservative 160;
```

```
Qy 32 ILKQTAQSSISGADYAESGSKLIKINETSFPV-----DDTVTLFSDKRTTPPKIKDN 86
Db 530 IASTLKAKPSMAGIDLSTYSTEKGEF---TAVPINLIDTDSITNI---RVVAVLEEDY 581
```

```
Qy 87 LAKPREQELKAVTENTESEKQITSGSQESKESLSINKTVP--STNWEICDFTTKGNT 145
Db 582 KANQONAKEVAYSVNA-----SOKTAVKVGRTVDMSDGNYFVFAISGNN- 628
Qy 146 LVGLSKSGVEKLSQTDHLVPSQAADGTQLIQVASFAFTPKKTAIAIAYTSRAGENGEBIS 205
Db 629 -----NQSDFTLATPVSNVAVVGLKSKPTVEFSLKEAQDKLT 666
Qy 206 QLDVDGKEIINEGVF--NSYLLKKVTIPT--GYKHIGQD-----AFVDNKNIAEV- 252
Db 667 -----INY-EVFDADNTLLYDNLNTHPTKLKYKSDAQGMVDSGPVATVDLENKOEIT 716
Qy 253 NLPE--SLETTISDYAFALHALKQID---LPDNLKAIAGELAFPDNOITGKLSLPROJ-- 303
Db 717 NLLEFGGLESTYVYVWVTASYNLDDGAGIMVDOJ--IGQSGVFQTKIAIKVATFLES 774
Qy 304 -----MLABRAFKN--HIKTIFRGNL-KVIGEASFQDNLSQMLPDGLEKI 351
Db 775 VTTNQAENVIKLSEAAKLNLSANLKIYDKKNTLVKTVP---LNDDFNKLMTEDGVTVL 830
Qy 352 ESEAFTPNGDDHNNRVLVMTKSGKNPGLATENTVYNDPKSLWQESPEIDYTKWLEED 411
Db 831 FEELAI---NKEYLVKVEDGLDSGMNQ--VPVEGQLVFYTK--KASPIITDKVLL---D 878
Qy 412 FTYOKNSVTGFSNKGLOKVKENKNLEIPKQHNGVTITE---IGD-NAFRNVDFQNKTLRK 467
Db 879 YOPNKLKVGGLA-----GIEATEKPLIDDYNAVSSI-----TVRI 913
Qy 468 YDLBEVLPSPTIRKIGA-----FAFOSNNL-KSFEASDDLEEKEGAFMNNRI--- 514
Db 914 YKVDDLTPLVEQEVSTAABEPERHSYFDLTNNALGRGFDYVYKAEVWINDYEDHRIEIS 973
Qy 515 -ETLELKDKLVITG-----DAAFHNHIIYAIVLP---ESVQIEIGRSAPRONGAN 559
Db 974 SETTIQKKEKTEVEYELRRTATEVQLNVVYRDIEGAIVPGKLEITSTGSGSEQLKNGKN 1033
Qy 560 NLIF-----MGSKVKTIGEMA---FLSNRLEHLDLSEQQLTETIPVQAQPSDN 603
Db 1034 TVTLPISSEGATTIKTTGDYVITAGEPAVTTDFDKTLVAMNTTAPKASAOI----- 1085
Qy 604 ALKEVLLPASLKTIREEAFKKNHLK---QLEVASALSHIAFNALDNDGDQDFD----- 654
Db 1086 ALDETGRLKITPEPDDVAKSNVMKTVTVSLQEAKSTSP---DYTEAKIGDQDFPQNLSL 1142
Qy 655 ---NKVVVKTTH-----NSYALADGEHFIVDPDKLSSTIVDELEKILKLEIG 697
Db 1143 PFGNVWFNKSHELTLDMKNYSERNMNEKLF--ENYVVSIDNSSAFVTSNGGLSTTRN 1200
Qy 698 LDYSTLRQTTQ-----TQFRDMTT-----AGKALLSKS-----NLRQGEKQK 734
Db 1201 LNRADVVRVTKGPTDSEGNIAGVKPKNIWTKYIAYRGVGMISNSETADPFSPFRQEDGS 1260
Qy 735 FLOE-----AQOFLGRVDLDKAIKAE-----KALVTKKAT-KNGQLL 771
Db 1261 YVAELNGRYVDFSGSLVSTEAGSKIDLYSAQEGVSQVSOAITTKALVEPNITAEVSIY 1320
Qy 772 ERSINKAVLA--YNNSAIKKANVKLEKELDLLTGLVEGKGLAQATWQGV---YLLKT 826
Db 1321 DRRVKLDVVGEDKONTTVKKN---ENEL-FVNAYKEDGETLVKSRVIDGLPTRIDFVT 1375
Qy 827 PL-PLPEYIYIGL-----NVYFDKSGKLIYALDMSDTIGEGOKDAVGNPILN 871
Db 1376 DLSPDKKYVIKVEGKYDLDGEGPQNKVYIYSETITTEKSLPSMSESTYSVSWFPAFY--ART 1433
Qy 872 VDENEGYPHALAVATLADYEGLDIKTILNSKLSQLTISIQ-----VPTAAY 917
Db 1434 IKGNNFIDESSVLNIEYRLYDASTI-NAELSDVALEQELGKTTPAATF 1483

RESULT 46
T13734
groovin gene protein - fruit fly (Drosophila melanogaster) (fragment)
```

Query Match	3.6%	Score 181.5	DB 2	Length 1751
Best Local Similarity	19.8%	Pred. No. 1.9		
Matches	187	Conservative 134	Mismatches 352	Indels 271
				Gaps 41
Qy	139	FITKGNLTVLGSLKSGVEKLSQTHLVLPSSQAADGTOLIQVASFAFTPKDKTAAIAEYTSRA	198	
Db	13	FVTKCQCETESYKQLVAKLDKLAIV-----DGYELFHKKKLG-----ENDIKVETNAS	62	
Qy	199	GENGEISQLDVGKKEIN-----EGEVFNSYLLKKVTI-PTGYKXHGIDQAFVDNK	247	
Db	63	ANNNNNQSVLTSKIRNFLSKFLQLPQHTDLLHLIRELAVEPNGIKYLVSEYEFNQ	122	
Qy	248	NIAEVN-----LPSLETTISDYAFAPHLAKQIDL-----	276	
Db	123	LMHVIHFHYDLLRAKLDHMCADHYCKIPEHLK-ISK-----KELDMUKVVLGYVRKP	173	
Qy	277	PDNLK-AIGELAPFF--DNOITGKLSLQPLRMRLAERAFKSNHIKTIEFRGNSLKVGEAS	333	
Db	174	LDNIKODIGKLETFITKNKTIKNISD---LIIAENKKRGSHPTTTTNGAGTQPANGSIA	230	
Qy	334	FQNDLSQLMLPGCLEKIESSEATGNGPDHYNNRVVLWTKSKNPSGLATENTYNPDK	393	
Db	231	AASSETTQ-----ISGSSNSGSSSTGSSNS-----GSSSTGSGTGTGQSPPA	276	
Qy	394	SLWQESPEIDYTKWLBEFTYQKNSVTGFSNKGLOKVKR-----NKNLEIPIKOHNGVTIT	448	
Db	277	AADASSTNANYE--AKKIIYQAVYNTIFVTNQLQEAQKLIADVLEKRVKVLKEHKDIKV-	332	
Qy	449	EIGNAFNRVDFQNKTLRKYDLBEVKLPSTIRKIGAFAGFOSNNLKSF--EASDDLBEIK-	505	
Db	333	-----LLEQVAKKEKLPSPDY-----NTTTLNVHKEASSKIALEK	370	
Qy	506	-----EGAFPMNRIETLE--LKDKLVTIGDAAPHNHHYIAIVLPSVQBOIGR	550	

```
Db 371 KIEAIKTVNFDLGLFTD---ABEEVYLREKAKWAG-----TLIIPSTYSAGT 418
Qy 551 S-----AFRQNGANNLI FMSKVKYTLGEM-----AFLSNRLE 582
Db 419 PGKTVPTLKETYPHGISVALAENSIYELIEKIGSDFTGDLQNPDDGKQPKKGLINETK 478
Qy 583 HLDLSEQQLTEIPVQAFSDNALKEVLLPASIKTIRERAFKKNHLKQLEVASALSIAFNP 642
Db 479 RKELLE-KIMNKIKIE-----BDKLP-NLKKEYEBKYVYEAKVNEPKPAFNHFYEA 528
Qy 643 ALDNDGDEQFNKVVVTHNSYALADGEHFI VDPDKLSSTIVDLKILKILGDLDTST 702
Db 529 RLDNTLVNKKFDD---FKKREAYMEKKKLESCSYEQNSNLINKLKGLTYLE---DVLV 583
Qy 703 LRQTQTQTFRDMTTAGKALLSK-----SNLRQGEKQKFLQBAQFPLGRVDLDKAIAKAEK 757
Db 584 RKDIADDEIKHFSFMWKLSKSEIYDLAQEIRKNENKLTIEKDFSGVVEL-----QVQK 638
Qy 758 ALVTYK--ATKNGQLLERSINKAVLAYNNSAIKKNVIRLEKELDLLTGLVEGKGPLAAQ 815
Db 639 VLIIRKIKALKNVQNL-----LKNKAVK-----MSDTIG 858
Qy 816 TMVQGVYLLK---TPLPLPEYVIGLVNVEDKSGKLIYALD-----MSDTIG 858
Db 662 ---DLYIIPKVKYTKSEKPEPYVLMVKREIDKLDKDFIPKISMIATEKKNKPTVAADIVA 718
Qy 859 EQQK-----DAYGNPI-----LNVYEDNEGYHALAV---ATLADYEGLDIKTILNSKLS 904
Db 719 KGSLRGASGTGTTGNTVNAQTAVVQPHQVQVNAVTVQPGITGHQAQGEAETQNS--V 776
Qy 905 QLTSTRQVPTAAYHRAGIFQAIQNAABAEQLLPKPGTHSEKSS 948
Db 777 QAAQVQQTQFAGAGQGVASTQTIQAPAPT-QASPEPAPAAPPST 819

RESULT 48
OVBX
adenylate cyclase (EC 4.6.1.1) - yeast (Saccharomyces cerevisiae)
N/Alternate names: ATP pyrophosphate-lyase; protein J1401; protein YJL005w
C/Species: Saccharomyces cerevisiae
C/Date: 28-Dec-1987 #sequence revision 08-Sep-1995 #text_change 09-Jul-2004
C/Accession: S56776; S56775; A24776; S05828; S55183
R/Rio Van, D.; Pereira, J.; Jacq, C.
submitted to the Protein Sequence Database, September 1995
A/Reference number: S56776
A/Accession: S56776
A/Molecule type: DNA
A/Residues: 1-1823 <DEH>
A/Cross-references: UNIPROT:P08678; EMBL:249280; GSPDB:GN000010; MIPS:YJL005w
R/Ride Haan, M.; Smits, P.H.M.; Grivell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A/Reference number: S56771
A/Accession: S56775
A/Molecule type: DNA
A/Residues: 673-2026 <ZAG>
A/Cross-references: EMBL:249280; GSPDB:GN000010; MIPS:YJL005w
R/Rataoka, T.; Broek, D.; Wigler, M.
Cell 43, 493-505, 1985
A/Title: DNA sequence and characterization of the S. cerevisiae gene encoding adenylate
A/Reference number: A24776; MUID:86079531; PMID:2934138
A/Accession: A24776
A/Molecule type: DNA
A/Residues: 1-261, 'L', 263-547, 'L', 549-591, 'H', 593-708, 'I', 710-961, 'P', 963-1387, 'S', 1389-
A/Cross-references: EMBL:M12057; NID:g171359; PIDN:AAA34549.1; PID:g171360
A/Note: the authors translated the codon TTA for residue 262 as Ser, ACG for residue 311
2 as Leu, TCA for residue 1566 as Ala, GAG for residue 1659 as Ala, GGT for residue 1739
R/Mason, P.; Lenzen, G.; Jacquemin, J.M.; Danchin, A.
Curr. Genet. 10, 343-352, 1986
A/Title: Yeast adenylate cyclase catalytic domain is carboxy terminal.
A/Reference number: S05828; MUID:86165073; PMID:3327602
A/Accession: S05828
A/Molecule type: DNA
```

```
A/Residues: 1042-1426, 'D', 1428-1460, 'T', 1462-1955, 'V', 1957-2008, 'M', 2014-2015, 'TNFYKWLRT'
A/Cross-references: EMBL:X03449; NID:g3487; PIDN:CAA27175.1; PID:g3488
R/de Haan, M.; Smits, P.H.M.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A/Reference number: S55183
A/Accession: S55183
A/Molecule type: DNA
A/Residues: 673-2026 <DEW>
A/Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60917.1; PID:g854568
C/Genetics:
A/Gene: SGD:CYR1; CDC35; MIPS:YJL005w
A/Cross-references: SGD:S0003542; MIPS:YJL005w
A/Map position: 10L
C/Superfamily: yeast adenylate cyclase; leucine-rich alpha-2-glycoprotein repeat homolog
C/Keywords: CAMP biosynthesis; duplication; phosphorus-oxygen lyase; tandem repeat
F/669-1343/Region: leucine-rich 23-residue repeats
F/1065-1087/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR>
F/1610-2026/Domain: catalytic #status predicted <CAT>
F/1664-1749/Domain: yeast adenylate cyclase catalytic domain homolog <YACC>

Query Match 3.6%; Score 181.5; DB 1; Length 2026;
Best Local Similarity 20.1%; Pred. No. 2.3;
Matches 223; Conservative 145; Mismatches 365; Indels 375; Gaps 48;

Qy 29 KEPILKQTOA-----SSSI---SCADYABSSGSKKLKINETSGPV----- 65
Db 616 KTPILNKTYSYTKFTSSVNMNPSDQAQSSG--LLIQDEKDDVRCQLEHYKDFSDLD 673
Qy 66 -----DQVTDLPDSKRTTPKIKDNLAKPREQELKAVTENTESEKQITSGS 113
Db 674 PKRYAIRFNPTDFTTLCTPATTVVEIIPALKR-----KNVITAQG 717
Qy 114 QLBSQKESLSLANKVTPSTSNWEICDFITKGNLTVGLSKS-----GVEKLSQT---DHL 163
Db 718 NFQISLVKGLSKILRPTSKPIIIE--RKLILLNGYRKSDPLHIMGTIEDLSFVFKFLPH 775
Qy 164 VLPQQAADGTQLIQVAFPTPDKKTAIAEYTSRAGENGESQLDDVDGKEIINEGEVNS 223
Db 776 VTPSH-----PTPEQQRIM-----RSEFVHVDLRNMDLTPPIIF-- 811
Qy 224 YLLKKVITPTGYKHIGODAFVDNKNIAEVLNPELETISDYAFALHAKQLDL----- 276
Db 812 -----YQTSIESLDVSNANIFLP--LEPI-ESSIKLLSLRWVNIARSKPPS 857
Qy 277 -----PDNLKAIGELAFEDNQITGKLSLPR-----QLMRLA 307
Db 858 NITKAYKLVSELQRLQNFIRKVPNSIMKLSNLTILNLCNELESIPAGFVELKNLQLDLS 917
Qy 308 ERAP-----KSNHIKTI-----EPRGNSLVIGESQD 336
Db 918 SNKFMHYPEVINYCTNLLQIDLSYNKIQSLPQSTKYLVLKLAOMNLSHKLNFIQDLS-EM 976
Qy 337 NDLSQLMLPDGLEKIEGAFTGNPNRRVLTGSKGNKPSGLATENTYVNPDKSL- 395
Db 977 TDLRTNLN-----RYNRISSIKTNASNLQNLFLTDNRISNPEDTLP 1017
Qy 396 -----WQESPEIDYTKWLEEDFTYQKNSVTGFNGK-----LQKVKNK 434
Db 1018 KLRALEIQENP---ITSISPKDF-YPKNMTSLTLNKAQLSSIPGELLTKLSFLEKLELNQ 1073
Qy 435 N--LEIPKQHNGVITTEIGDNAPFNVDP-----ONKTLRKVDLEEVKLPSTIRKIGFA 486
Db 1074 NNLTRLQPEISKLVKLVLSVARNKLEYIPPELSQLKSLRTLDL----- 1117
Qy 487 FQSNLAKSF-EASDDLE-----EIKEGAFMNNRIETLELKDVLVTIGDAAPHNNIYAVL 541
Db 1118 -HSSNIRDVFGMENLELTSLNISNAFGSSLEN-----SFYHNSYG--- 1160
Qy 542 PESVOETGRSAFRQNGANNLI-----FMGSKVTKLGEMAFPLSNRLEHLDLSEKQ 591
Db 1161 ----SKLSKSLMFFPIAADNQPDAMWPLNCFVNLKVLNLSYNNP--SDVSHMKL---ES 1211
Qy 592 LTEIPVQAQSDNAL-----KEVLLPASIKTIREAPKKNHLKQLEVASALSH----- 638
```

Db 1212 ITL-----YLSGNKLTLSGDTVLKWSLKT-----MLNSNQMLSLPAELSLSLSQSVFD 1263
Qy 639 IAFNALDDNDGEQED-----NKVVVKTHNSYALADGEHFI---VDPDKLSSTIVDLK 690
Db 1264 VGANQLKYNISYHNDWNRNKKELKYLNFSGRRPEIKSFISHDIDADLSLTVLPLQK 1323
Qy 691 ILKLJEGLDYSTLRQTTQTFQFRDMMTAGKALLSKNLRQGEKQKFLQEAQFGLGRVDLDK 750
Db 1324 VLGLMDVTLNTTKVPDENNVFRLRTTASII-----NGMRGVADTLGQRDVSSRDVTFER 1379
Qy 751 ATAKAEKALVTKATNG-----OLLER----- 773
Db 1380 FRGNDDECLLCHDSKNQADYCHNISRIVRDIYDKILIRQLERYGDETDNDIKTALRFS 1439
Qy 774 --SINKAVLAYNSALKKANVRLEKELDLTG-----LVGK-----GPLAQATM 817
Db 1440 FLOLKEINGMLNSVDNGADVANL--SYADLLSGACSTVIYIRGKKLFAANLGDGMALISK 1498
Qy 818 VQGVYLLKTPPLP-----EY-YIGLNVYFDKSKLIYALDMSDTIG--EGQKDAYGNPIL 870
Db 1499 NNGDYQTLTKOHLPTKREYERIRISGGYVNGKLDGVDSRAVGFPDPLPHIASPDI 1558
Qy 871 NVDENEGHALAVATLADYEGLDIKTI 898
Db 1559 SVVTLTKADEMLIVATHKLWEYMDVDTV 1586
RESULT 49
AGI085
hypotheical protein lmo0086 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AGI085
R;Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunat, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AGI085
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1959 <GLA>
A:Cross-references: UNIPROT:Q8YAN1; GB:NC_003210; PIDN:CAC98301.1; PID:gl6409445; GSPDB:
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo0086
Query Match 3.6%; Score 180.5; DB 2; Length 1959;
Best Local Similarity 19.2%; Pred. No. 2.5;
Matches 206; Conservative 161; Mismatches 384; Indels 321; Gaps 52;
Qy 20 HNDGEVFLVKPILKQTOQASSISGADYAESGSKKLKINETSQPVDDTV--TDLFSDKR 77
Db 199 NNHYFDLKKEMLVQAKETDID-----IRSIKVEIDDAEKRSLKTAKE 244
Qy 78 TTFEIKONLAKPREQELKAVTENTESEKQI--TSGSQLEQSKS--SLSLNKTVPSTGNW 134
Db 245 TEKETTKDSTEK-----ESEKQNDASGQKSAQNNESESKDQOTQDAS-- 288
Qy 135 EICDFITKNTVLGSKSVEKLSQTDHLVPSQAADGTQLIQVASFAPTPDKTAIAEY 194
Db 289 -----GNNTNGSANNNTARSNGTGG---SNSDTNGSQ----- 317
Qy 195 TSRAGENISQLDVD-GKEII--NEGVEFNSYLLKKVTIPTYGKHIGQDAFVNKNIA 250
Db 318 --PGGTQKPNBGVDVKNANEIILKUNEAENTNTP-----QVP-----IV 354
Qy 251 EYNLPESLETISDYAFALHALKQIDLPDLNKAIG--ELAFPD--NOITGKLSLPLQMLRL 306

Db 355 DVNL-----TVK-----GQVAAAKLKLTDSSSKRLNSLEAILYDSKNIVKKEKLN---STK 402
Qy 307 AERAFKSNHIKTIIBRGNSLKVIGEASQDNDLSQLMLPDGLEKIESEAFTCNPGDDHVN 366
Db 403 ANQNPFTNNLK-----YGETYQVVQGSYK----- 427
Qy 367 NRVLWMTKSGKRNPSGLATENTYVNPDKSLWQESPEIDYTKMLEEDFTYQKNSVT-----G 421
Db 428 -----SASNKNQDIIFFQTV-----BAKPVVLTPKVTGERCEYMAELTATELYG 473
Qy 422 FSNKGLQKVKKNKLEIFKQHNGVTIIEIGNAFNVPFQNKTLRK---YDLEEVKLPST 478
Db 474 HIDKLVLKIKENNSVITSQKVTVDASQTKDQGVKFNLSDDKEYIIEEELIVDCK 533
Qy 479 IRKIGAPAFQNNLKSPEASDLE--BIKEGAFNNRIETLELKDVLVTIGDAA----- 530
Db 534 NVTDDGTFISLTAKKPTIEGLNLSYSTDGEFVSPNVLVDRDESITSRYVAYLEDD 593
Qy 531 FHIN-----HIYAIV--LPESVQIEGRSAFRQNG-----AN 559
Db 594 YKVGNSNAKEYAVSVVDANQKTA/KVGRITVDMNDGNVIFVYISGNNQSDYTFATPAS 653
Qy 560 NLIPIWGSKVTLGEMAFLSNRLEHLDLSEQQLTEIPVQAF--SDNALK--EVLLPASLTKTI 617
Db 654 NSVVVGKTKTPTVEFS-----LKEAEODKLT--INVEVFDADNTLLFDNLTHPTLKLY 704
Qy 618 REEA---PKKHHLKOLEVASALSALSHAFNALDNDGDEQFNKVVVKTTHNSYALADGEHF 674
Db 705 KSDAQGYSGNPNVATVDTLKSD---ITNLEFDGLSEAYYVVVMT--GSNLDGAGI 759
Qy 675 IVDPKLSSTIVDLKILKJIEGLDYSLRQTTQTFQFRDMMTAGKALLSKSLNR----- 728
Db 760 MVDELIGOSSVPRITEITKVNASPSLDSV-DTTKVAINVKLSDRAVNLNDANLKIYEKT 818
Qy 729 -----QOEKOKFLOEA--QPFGLGRVDLDKA-IAKAEKALVT--KKATKNGQLLERS 774
Db 819 NTLVKTVPLHGDPEKLMDSGNSYLFENLSINKEYLVKVEDGYDSGMNHPVPEGQFIKT 878
Qy 775 -----INKAVLAYNSAIKKANVRLEKELDLTLGLVEGKGL-----AQATWQGVYL 823
Db 879 KKEPSTDVKLLDY-----KASEMKLGGLAGIAATEKPLDIDYNAVSSITYSIYK 928
Qy 824 ---LKTPLPLPEYITGL---NVYFDKSGKLIYALDMSDTIGEGOKDAYGNPILNVD--- 873
Db 929 ADDLNTPLVEQEVSTAADFKNVFDLTNKL-----GRGYSY---VIKADVW 974
Qy 874 EDNEGHALAVATLADYEGLDIKTLNLSKLSQTSIRQVPTAAHY--RAGIFQAIONAAA 931
Db 975 NDNYEDHHIEISS---DTIQIK-----KEKPTVEYBILSRASEIKLVVYV 1017
Qy 932 EAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGHRAILPRTSGKS 983
Db 1018 EDEESIVPGTLEITSTTG-----GNEQLQSGKNNVTLP--LSSEGT 1057
RESULT 50
S02771
myosin heavy chain A [similarity] - Caenorhabditis elegans
N:Contains: myosin Arpase (BC 3.6.4.1)
C:Species: Caenorhabditis elegans
C:Date: 31-Dec-1993 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C:Accession: T23622; S02771
R;Harris, B.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19773
A:Accession: T23622
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1992 <WIL>
A:Cross-references: UNIPROT:P12844; EMBL:Z78199; PIDN:CAB01576.1; GSPDB:GN00023; CESP:KI
R;Dibb, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.
J. Mol. Biol. 205, 603-613, 1989

A>Title: Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy chain gene
A:Reference number: 802771; MUID:89178677; PMID:2925820

A:Accession: S02771

A>Status: nucleic acid sequence not shown

A:Residues: 1-116,140-1992 <DIB>

A:Molecule type: DNA

A:Cross-references: EMBL:X08067; NID:g6798; PIDN:CAA30856.1; PID:g6799

C:Genetics:

A:Gene: myo-3; CBSP:K12F2.1

A:Map position: 5

A:Introns: 46/1; 192/1; 292/1; 468/2; 1921/3

C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle co

F:89-802/Domain: myosin motor domain homology <MMOT>

F:202-209/Region: nucleotide-binding motif A (P-loop)

F:690-712/Region: actin binding #status predicted

F:793-807/Region: actin binding #status predicted

F:875-1992/Domain: coiled coil #status predicted <COI>

F:875-1189/Region: S2

F:1190-1992/Region: light meromyosin

F:153/Modified site: N6,N6-trimethyllysine (Lys) #status predicted

F:208/Binding site: ATP (Lys) #status predicted

F:730,740/Active site: Cys #status predicted

Query Match 3.6%; Score 180.5; DB 1; Length 1992;

Best Local Similarity 20.2%; Pred. No. 2.5; Mismatches 354; Indels 237; Gaps 43;

Matches 187; Conservative 149; Mismatches 354; Indels 237; Gaps 43;

2 KKHATVLTITVSVVTHNOE--VFSLVKEPILKQTQASSISGADYAESSGSKSLKIN 59

987 KKHQDLESLRKARQEQSRDHNRSL-QDEMANQDEAVAKLN--KEKHQESNRKLN 1043

60 ETSQGVDDVTVDLFSDKRTTPEKIKDNLAQPREQELKAVTENTSEKQITSGSOLESK 119

1044 EDLASEEDKVNLH-----EKIRNKL-----EQMDLEBENIDREKR--SRGDIEKAK 1088

120 ELSLNTKVPSTSNWEICDFITKGNLTVLGSKSGVE-----KLSQTDHLV 164

1089 KRVGDLKVAQ-----ENIDEITK-----QKHVDVETTLKKBEDLHHTNAKLAENNSII 1137

165 -----LPSQAADGTQLIQV-----ASPAFTPKKTAIAEYTSRAGENG--EISQL 207

1138 AKLQRLKELTARNALAELEAEARNRQKSDRSSEARELEELTERLEQGGATAQAL 1197

208 DVDGKEIINEGEVP-----NSYLLKKVTIPTGYKHIGQAFVDNKNIAEVNLPSELET 260

1198 EANKK---REABIAKLREKEBEDSLNHETAISSLKRHGD-----SVAB--LTEQLET 1245

261 ISDYAFALHALK---QIDLPNLKAIAGELAFPDNQITCKLSIPROLMLAERAFKSNHIK 317

1246 LQKLKAKSEAKSKLQRLDERSQHAT-----DSEVRSRQDLEKAL-----K 1286

318 TIEFRGNSLKV-IGBASPDNDLSQLM-----LPDGLEKIESEA----- 355

1287 TIEVQYSELQTKADQSRLQDFAALKNRLNNSDNRSLSEMDNQLSLHLRLKSTLQS 1346

356 ---FTGNGPDHNNRVVLTWTKSGKNGPGLATENTYVNPDKSLMOESPEIDYTKWLEBDF 412

1347 QLDERTNRYDEESRQAL-AATAKN---LEHENTILREHLDEEAES-KADLTRQISKLN 1401

413 TYQKNSVTGFGNKGLQVKR---NKNLEIPKQHNQVITTEIGDNAPRVNDFQNKTLRK- 467

1402 AEIQWKAQFQSEGLNKLEETEAAKALQLKQVE-----LTDNGLFAKIASQEKVRPKL 1457

468 -YDLEVKLPSTIRKIGA-FAFQSNLKSFEA-----SDDLLEEIKEGAFMNNRIETL 517

1458 MODLDDAQ--SDVEKAAQVAFYEKRRQFESIIAEMKKTKDLSSELDAQQRNQLST 1515

518 EL-----KDKLVTTGDAAFHINHIYIVLPESVQEIQRSPQNGANNLIFMGSKVITL 571

1516 DLFAKTAANDLAEYLDSTRENRKSLAQEVKDLTDQLGEGG-----RSV 1559

572 GEMAFPSNRLEHLDSLSEKQLTEIPVQAPSDNALKEVLLPASLTKTIR-----BEA 621

Db 1560 AELQIVRKLEVEKEEELQKALDEARALAEBAEA-KVTRAQIEVSQIRSEIEKRIQEKEEE 1618

Qy 622 F---KXNHLKQLEVASALSIAF-----NALDDNDGDQFONKVVVKTTHNSYALADGEHF 674

Db 1619 PENTRRNHQRALESQMATLEAETKQKEALRIKKLESDDINDLBIALDHANRAYADAQKT 1678

Qy 675 IVDPDKLSSITVDLEKILKIEGLDYSTLRQTTQ-TOPRMTTGTAGKALLSKNSRQGEKQ 733

Db 1679 I-----KKYMETVQELQFQIEBEQOKQIREQFLASE---KFNAILQSEKD 1722

Qy 734 KFLQEAQFFLGRVLDLKAIAKAE-----KALVTKKATKNGOL-----LE 772

Db 1723 ELAQQA-----AARARNRAEACIELRBQNDLNHAUSALTGQRKLEGLLAHAHELE 1778

Qy 773 RSINKAVLAYNNSAIKKANVKRLEKEL 799

Db 1779 EIANELKNAVEQGGKASADAARLAABEL 1805

RESULT 51

JC6009

surface-located membrane protein Imp3 precursor - Mycoplasma hominis

C:Species: Mycoplasma hominis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: JC6009

R:Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.

J. Bacteriol. 178, 2775-2784, 1996

A>Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system

A:Reference number: JC6009; MUID:95213016; PMID:8631664

A:Accession: JC6009

A:Molecule type: DNA

A:Residues: 1-1302 <LAD>

A:Cross-references: UNIPROT:Q49547; EMBL:X95601; NID:g1197335; PIDN:CAA64858.1; PID:g119

C:Genetics:

A:Gene: lmp3

A:Genetic code: SGC3

C:Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homology

C:Keywords: duplication; membrane protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-1302/Product: surface-located membrane protein lmp3 #status predicted <MAT>

F:957-992/Domain: tetratricopeptide repeat homology <TT1>

F:983-1026/Domain: tetratricopeptide repeat homology <TT2>

F:1089-1120/Domain: tetratricopeptide repeat homology <TT3>

F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match 3.5%; Score 180; DB 1; Length 1302;

Best Local Similarity 18.6%; Pred. No. 1.4;

Matches 199; Conservative 183; Mismatches 417; Indels 269; Gaps 46;

Qy 40 SSISGADYAESSGSKSLKIN--ETSGPVD---DTVTDLFSDKRTTPBKIK-----DNLAK 89

Db 125 NSNDQGVNDSDAKALNENQIDDSLPIDIKTKTNENLNAKELLNKAERELQSKIF 184

Qy 90 GPREQLKAV--TENTES-----EKQITSGSOLESKESL-----SLNKTV 128

Db 185 NEKKQELKRVLDLDETEVDFTKEQKVFIEINNETSIEDIKNKILIEVEKATSSLSKI 244

Qy 129 PSTSNWEICDFITKGNLTVLGSKSGVEKLSQTDHLVPSQAADGTQLTQVASFAFTPKK 188

Db 245 LNTQQLQELQEF---ENIKKQLQDFINTKLDANKAQYSIKQKALDKINSINGI-----NKN 295

Qy 189 TAIAYTSRAGENEIISQLD-----VVG-----KEIINEGEVFNPSVLLKKVTIPT 233

Db 296 STIKEL--KAGQNALIKAKGEAGLEKEKLDQGNIKDKTLKETINNAKEP-----KKLLIDN 348

Qy 234 GYKHIGQDAFYDNK-NIAEVNL---PESLETISDYAFALHALKQIDLPDLNKAIGELAFF 289

Db 349 DQKITVDLKSJLNDNEISKAEQSLSKDKESMESANOL-----LNTKLIYEKILINKFNQ----- 400

Qy 290 DNOITGKLSLPLQRLMLAERAFKSNHIKTIFFRNSLKVIGEASFQNDLSQLM--LPDG 347

Db 401 -----EKEAKFNELE--QTRQNIENFLTDEVKNNPNVATLVKDLTNA 440


```
Db 854 EAVSVQRSMQETVWKLHQEQFN---MLSSDLKRLNADMEAKFREKDEREQLIK 910
QY 439 PKOHNGVYITEI---GDNAFRNVDFQNK-TLRKYDLEEVKLPSTIRKIGAFQSNLKL 493
Db 911 AKEKLENDIAEIMKSGDINSQUTKWNDELRKEDVEELQKLT-----KANENA 961
QY 494 SP-EASDDL-----EIKEGAFMNNRIETLEKOKLVTIGDAAFHINHIYAIVLPSVOE 547
Db 962 SFLQKSIEDMTVKAEOQQAEEAKKHEEKELERKLSL-----EKKME 1005
QY 548 IGRSAFRONGANNLIFWGSVKYTLGEMAFLSNRLEHLDLSEQKQLTEIPVQAFSDNALKE 607
Db 1006 TSHNOQOE-----LKARYERATSETKTHEEILONLQKLLDTE----- 1044
QY 608 VLLPASLKTIRBEAFKKNHLKOLEVASALSASHAFNALDDNDGDEQPDNKKVVKTHNSYA 667
Db 1045 -----DKLKGAREE---NSGLLOLEELRKQADKAKAQAOTADAMQINQWTKTE---T 1094
QY 668 LADGEHFI VDPDKLSSTIVDL-EKILKLI EGLDYS-----TLRQTQTQFR-DMTTAGKAL 721
Db 1095 LASLEDTKQNAKLQELDTLKENNLKNVBEELNKSKELLTVENQKMEFRKEIETLKQAA 1154
QY 722 LSKS---NLROEKQKFLQEAQFFLGRVLDLKAIAKAEKALYTKATKNGQLLE----- 772
Db 1155 AOKSOOLSALQENVKLAEE---LGR---SRDEVTSKQLEERSVLNQLLEMKKRES 1207
QY 773 -----RSINKAVLAYNNSAI---KCANVRLKELDLTLGLVECKGFLAQATWVQ 819
Db 1208 KEIKDADEKASLOKSISI---TSALLTEKDAELEKURNEVTVLRG---ENASAKLSHVQ 1263
QY 820 GYVLLKTPLEPYEYIGLVNVPDK-----SGKLIYALDMSDTIGEGQKDAYGNPLNV 873
Db 1264 TLSDKVKLELVKQNLQLEKQKRLSSSGNTDTQADEDERAQESQIDFLNSVVDLQ 1323
QY 874 EDNE 877
Db 1324 RKNQ 1327

RESULT 55
H71611
probable secreted protein PF80565w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C:Accession: H71611
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Pertes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: H71611
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1817 <GAR>
A:Cross-references: UNIPROT:096205; GB:AE001403; GB:AE001362; NID:g3845216; PIDN:AAC7190
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF80565w
```

```
Query Match 3.5%; Score 177.5; DB 2; Length 1817;
Best Local Similarity 18.7%; Pred. No. 3;
Matches 199; Conservative 159; Mismatches 394; Indels 315; Gaps 48;

QY 16 SVYTHNQEVPSLVKPEILKQTOASS-----ISGADVAESSG-----KSKLKINE 60
Db 577 NYEQOKQKQTLVYINDKIKKKNKNSYEMNNNTYMHGTYTDIENRLNKKKRLNVRG 636
QY 61 TSGPVDVDTDLFSPDKRTPEKIKONLAGPREQ---ELKAVTENSEKQITSGSQL- 115
Db 637 RTNTLDDIIV---SDHNSYDKY---NTSKHNRKKNHINEMKKQKNNKQNTLFLVDGKME 691
QY 116 ----EQSKESLSLNKTVFSTSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQAAD 171
```

```
Db 692 GIGKEKENKMNNN-----IFYNNSYNINNNSSYNINNDIYSDNMTSVN 739
QY 172 GTQLIQ-VASFAFTPDKTAIAEYTSRAGENCEISQOLDVDGKEIINEGVFNSYLLKKYT 230
Db 740 NTKYVSGVPVSAHLINKQVNEYQQLPNYN-----NMMIKGSHIINELPK-NNYIYEN-- 792
QY 231 IPTGKHHIQDQAFVNDKNIABVNLPESLETISDYAPAHALAKOIDLDPNLKAIGELAPFD 290
Db 793 -----NYIGQVYLTNP-LYNKETKDIFYIYKYL-----KIIYSPSLKK---RMEFID 838
QY 291 NOITOKLSLPROLMRLAERAFKSNHIKTIEFRGNLSKVIGEASFOQNDLSQLMLPDGLEK 350
Db 839 NCWKTKIFVIRKVCNFKRPFSSNK-KNNKMNDRSDSYVDNISSYYDDD----- 885
QY 351 IESFAFTGNPGDDHNNRVNLWTGSKNPSGL-----ATENTYVNPDK 393
Db 886 -----NNNNNNINILKCKKKKREVGGGIRLNGVDNKRKTHDDTIDIEKYKNNRN 935
QY 394 SIWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVKRNKNLEIPKQHNGVTITEIGN 453
Db 936 YLFNMGVDVLYNK---DQLGYKNS-----LDDN 961
QY 454 AFRNVDFQNKTLRYDLEEVKLPSTIRKIGAFQSNLKSFEASDDLEIEKEGAFMNNR 513
Db 962 NNNNNYNNNDNIRRHVSVC-----SYRRAHNIK-----YDIKEGGSNDIY 1003
QY 514 IETLEKDKLVITIGDAAFHINHIY-----AIVLPESVQETGRSAFRONGANN 560
Db 1004 TSNIKRNKKTKNTIEEINFINSMNLNKEALKNYVTVDTILYDEFSKLLKGIFEK---NK 1060
QY 561 LIFMGSKVKTLCGEMAFLSNRLEHLDLSBOKO-----LTEIPVQAFSD- 602
Db 1061 CLF-----KLKENYWSQNSYLLHLKDIKCRCTCLNYQRLLFHEVINLFVYVYKFCNW 1114
QY 603 NALK---EVLPLASLKTIREBAFKKNHLKOLEVASALSASHIAPNALDDNDGDFQDNKV 659
Db 1115 DVLKNYFDILINGS-----EEA-----IIVLEHFR-----NINKEQID--VIR 1151
QY 660 KTHNSYA-LADGEHFI VDP-----DKLS-----STIVDLKILKLEGLDYSTLR 704
Db 1152 KSYNNMYEYLSKSKYEHIIDDIINDYNNKINNMERKINIRIIDIIFK-----EYLLLI 1206
QY 705 QT---TQTFQDMMTAGKALLSKNLROGEKQKFLQEAQFFLGRVLDLKAIAKAKALVT 761
Db 1207 QOEHTKEGLKKNHIYKSKILFKNPLPSNLLKI-----ILCDKKEKIKENLNTN 1257
QY 762 KKATKNGQL-----LERSINKA-----VLAYNNSAIKKAN 791
Db 1258 CFSTVNNMLRNDMIKSGTFYFSKYSYCFERLLDYAFSLTITFENINFIINYIGDVLKLYE 1317
QY 792 VKRLEKELDILTLGLVEGK--GPLAQATWQGVYLLKTPLEPYEYIGLVNVPDKSGKLIY 849
Db 1318 VD-FKNSLYLLVITIKLHKFINNLPFEITKVREI--LTKVSI-----LKNKYVERI-KLLY 1369
QY 850 ALDMSDTTIGEQKDAYGNPILNVDEDEGNYHALAVATLADYEGLDIKTLNLSKLSQTSI 909
Db 1370 LLNCIPLI-----YLDPHNVLINESY-----EYKMDDDDKIIFSKLSPFSLV 1413
QY 910 RQVPTAAVYHRAGIFQAIQNAAAAEQLPKPGTHSEKSSSSSESANSK 956
Db 1414 SKV---VNHK---LRSVYTYHDYTDNLNLENEPFIHKNKTSKSMNDDTK 1454

RESULT 56
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676; A45521
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
```

RESULT 57					
T05113	hypothetical protein F28M20.240 - Arabidopsis thaliana				
C:Species:	Arabidopsis thaliana (mouse-ear cress)				
C>Date:	23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004				
C:Accession:	T05113				
F:By:	Bevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Hoheisel, J.				
A:	submitted to the Protein Sequence Database, November 1998				
A:Reference number:	Z15398				
A:Accession:	T05113				
A:Molecule type:	DNA				
A:Residues:	1-2712 <REV>				
A:Cross-references:	UNIPROT:Q9SB74; EMBL:AL031004				
A:Experimental source:	cultivar Columbia; BAC clone F28M20				
C:Genetics:					
A:Map position:	4				
A:Introns:	17/3; 240/1; 1950/1; 2118/3; 2381/1; 2599/3; 2645/1; 2679/2				
A>Note:	F28M20.240				
Query Match	3.5%; Score 177; DB 2; Length 2712;				
Best Local Similarity	19.7%; Pred. No. 5.8;				
Matches	220; Conservative 177; Mismatches 373; Indels 346; Gaps 54;				
Qy	20	NNQEV-----FSLVKEPILKQTQASSISGADYAESGSKSL-----	56		
Dd	462	HNOFVAEISQLRASYSAAVTNRNDSLAEBSCKSLYAATSSNTLNLENOLLATEAQVEDF	521		
Qy	57	--KINETSQGPVDVTDLPSDKRTTPPEKIKDNLAKPREQEELKAVTENTSESKQITSGSQ	114		
Dd	522	TAKNELQLSLEKSLDLLSETKKFNQLVEN-----DTLVAVISSMNDEKK-----E	569		
Qy	115	LEOSKSLSLUNKTVPTSTNNWEICDFITKGNTLVGLSKSGVEKLQSOTDHLVLPSQAADGTQ	174		
Dd	570	LIEBKE--SKNYBIKHLS--ELCNC---KNLAAILKAEEVPENT---IGP-----	611		
Qy	175	LIQVSAFPDPKKTATAETSYTSRAGENGEISQLDVQGEIINGREVPNSVLLKKVIPTG	234		
Dd	612	-----LTDEKIHLVVEEKYSLLGEAEKLQEELANCKTVVTTLQEVENSMMKETLSLLTR	663		
Qy	235	YKHIGODAFVDDNKNIARVNLPESLE-----TISDYAFAHLAKQ-IDLPDN--LK	281		
Dd	664	QQTMFER--NNTHLRENEKAHLELSAHULISETYLLSEYS----NLKEYTILLNKKLLK	716		
Qy	282	AIGELAFPNDNQITGKLSLPQLMLRAERAPKSNHIKITPIRGNSLKV-----	328		
Dd	717	FQGEK---EHLVENDKILTQELLTLQE-----HMSTVBEERTHVELREAIRLDKLA	767		
Qy	329	-----IGEASFQDN---DLSQLMLPDGLEKIESEAFGNPCDDHYNNRVVLWTK	374		
Dd	768	EENTSLSSTMVKARMVDNGSADVGLINQEIASEKLGSRSEIG-----SVTG	812		
Qy	375	SGKNPSGLATENT--VYNPPDKSLMQESPEDITYTWKBEDFTYQKN-----	421		

Db 813 SKQSAFL--ENTQYTNLBEVREYTSFSEFALMKNLEKEMQVONLEEAIKQILDTSSVSK 870
QY 422 FSNKG-----LOKVRKNKLEIPKQHNGVTITEIGD--NAPRVNDFQNTKLRKYDLEE 472
Db 871 SDKGATPAVKLIQAFESKRPEPESENAQITDLDSEADQFVSVNQIRNLRL-LDQ 929
QY 473 VKLPSTIRKIGAPAFQSNLKSFEASDDLEEIKEGAFMNRITLLEKDKLVITIGDAAPH 532
Db 930 LLL--NARKAG--IQFNQNDRTS-----TNQRLLELV-----EFASH 965
QY 533 INHIYAIVLPESVQIEGRSAFR-----QNGANNLIFMGSKVKTIGE-----M 574
Db 966 QOHINVLEADTIESVFEALXHYSELOHKNHLELLCDSLKLRNDNISVENTELNKKL 1025
QY 575 AFLSNRLEHLDL---SEQOLTEIPQAFSDNALKEVLLPASLKTIRBEAFKKNLQLE 631
Db 1026 NYC/SURIDELETOENLQNLTSFL-----STMEEQLV--ALQDESERAMVEH-----E 1073
QY 632 VASALSHI--AFNALDD---NDGDEQFONKVVVKTTHNSYALADGEHFIVDPDKLSSTIV 686
Db 1074 LTLMSFEGEAVVRLDDCLLRS-----TSGAHTGLDMTKRISGSV 1114
QY 687 DLKILKLEGL-----DYSILRQTTQTFR-----DMT- 715
Db 1115 DV--AVNVIEDLKELEAAVYKHSTSNKYBELKOSFNTLFEKNEFTASSMOKVYADLT 1172
QY 716 ---TAGKALLSKNL-----ROGEKQFLOEAOFFLG-RVDLDKAIKAKALVT 761
Db 1173 LITSCGSAMTSLEVENVAFDPFDGSEFENLEAVRILSERLELQSVIDKLOSL-S 1231
QY 762 KKATNGQLLERSINKAVLAYNNSAIKKANVRLEKELDLTLGLVEGKGLAQATMVQGV 821
Db 1232 SKSDMEEMTQSLDSTSL-----RELVEKVEGLELESV-----1268
QY 822 YLLKTPLPPEYVIGLVNFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDNEDNEGTHA 881
Db 1269 ---FESPSSQVEFLVSQLV-----KFIBTEELANLIRK-QLEAKGNELMEI-EBSLLHHK 1319
QY 882 LAVA-----TLADYEGLDIKTLNLSKLSQTSIRQ-----VPTAAYHRAGIF--Q 924
Db 1320 TKIAGRESLQARESLVAVRSELQKSNELSQSORLLSTREKUSIAVTKGKGLIVORD 1379
QY 925 AIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGL 960
Db 1380 NVKQSLAEASAKLOK-----CSEELNSKDARL 1406
RESULT 58
AC1583
internalin protein (LPXTG motif) homolog lin1204 [imported] - Listeria innocua (strain O
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
R:Accession: AC1583
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, P.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1583
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1237 <GLA>
A:Cross-references: UNIPROT:Q92CG8; GB:AL592022; PIDN:CAC96435.1; PID:g16413678; GSPDB:C
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin1204
Query Match 3.5%; Score 176.5; DB 2; Length 1237;
Best Local Similarity 20.1%; Pred. No. 1.9;
Matches 226; Conservative 179; Mismatches 365; Indels 353; Gaps 64;

QY 1 MKKHLKTVL-----TLTTSVVVTHNQEVFSVLKBEPIKQTOASSSISGADYAESGK 53
Db 1 MKRKISSIIVVIGIMLFQSLTWPFYITAKE--NEQKEEINKPSKITKGLT-----48
QY 54 SKLKINETSQVDDTVTDLPSDKRTTPPKIKDNILAKGPREBELKAVTENTSEKQITSGS 113
Db 49 NSLYTKTILETGTDTYDSVFPDSALA--KVVAKEATG-----SENT--TOLVTQA 94
QY 114 QLEQSK-----SLSINKTVPSTSNWEICDF--ITKGNLTAVGLSKSGVEKL 157
Db 95 DLNKIKSLNGYNGKISVLTGIDLLVNVTSISLNNQVTDISPIDQLPNLVSLS-----VKN 150
QY 158 SQTDLHLVPSQ-----AADGTOLIOVASFAFTPKK--TAIAEY 194
Db 151 NOISSLILNAQNLPKLTIDIENNPDLNTIDIQDPOLVDVKTSGYTLKRLTIVIAK- 209
QY 195 TSRAENGESIQDVGKKEINEGEVFNYSLLKVTIPTYGYKHIGQDAFVNKNIAFENL 254
Db 210 -----NNPELVNLGQ--YTIIRNV-----YFSQVASTKVEL--VNL 241
QY 255 PE-----SLE-----TISDYAFAPHALKQIDLPDNLKAIGELAFDFDNOITCKLSLPRQL 303
Db 242 PKVRKVNLRNSINELKVTDLAIEDLPLEENELDTV-----FDN-----IQNLP---286
QY 304 MRLAERAFKSHIKTIEFRGNSLVIGEASPDNDLSQMLPDLGLEKTESAFTGNPGDD 363
Db 287 -----NLKTLDSKNQLEEVLDKTDVENLPNLMTLIDIOQNLAKILI--NVODQ 333
QY 364 HYNRVVLTWTKSGKNPGLATENTVYNDP-KSLWQESPEIDYTKWLEBDFYQKNSVTGF 422
Db 334 --POLVDVKTSDYKELSALTIVIAKNPELVNLGQ-----YT--IRNVYFSQVASTKVK 383
QY 423 SNKGLOKVRKNLE-----IPKOHNGVTITEIGDANFRV-----DFQN 462
Db 384 ELVNLPKV-RKVNLRNSINELKVTDLAIEDLPLEEN-----ELTDTVFDNIQNLPLKT 437
QY 463 KTLRKYLEEVLKLPST-----IRKIGAFQSN--NLKSFASDDLEIEKEGAFNV-NRI 514
Db 438 LDLSKNQLEEVLDKTDVENLPNLMTLIDIOQNLAKILNVODPOLVDVKTSDYKELSAL 497
QY 515 ETELEKO--KLVTIG-----DAAPHINHIYAIVLPE-SVQEIGRASFQNGANNLI 562
Db 498 TTVIAKNPELVNLGYPIMQNVYFSQVASTKVELANLPKRAVRLRNSINQIELNVLV 557
QY 563 -----FMGSKVKTIGEMAFLSNRLEHLDLSEQ-KQLEIPIQAFSD 602
Db 558 SVKDVNLTNKTINDSIEKFRGMPI--LATNLNKNQITNINMLDGPFPQMTLNLIDL---612
QY 603 NALKEVLLPASLKTIREAFKKNHLKQ-LEVASALSHTAFNALDDNDGDEQFDNKKVVK 661
Db 613 NSIS--VLPSNLKTKPKLSISALNQVTIGKAI-----VVDSD-----LIINN 656
QY 662 HHNSVALADGHEFIVDDPKLSS-----TIVDLKILKLE-----GLDYS 701
Db 657 EISNF-----EKLTEPSPISNFGTYANEKITWSSERIKNLTEVSQFSQSELINVTGIDGT 710
QY 702 TLRQTTQTFQDMTTAGKALLSKSNLROGEKQ---KFLQEAQFFLGRVLDLKAIAKAKA 758
Db 711 FSGKVQ-PFKRSTTPVFNADSEIHYPQGTETKAEFLKDVQ---AQTDDLAIKASDFET 766
QY 759 LVTKK-----ATKNGQLLE--RSINKAVLAYNNSAIKKANVK-RLEKEL-----DLI 802
Db 767 MVNLKVKGYTVTLNVENLDGIKATPKKEIVVID-AVQGANITVKYDESGNKLAEISIL 825
QY 803 TGLV-----EGKGLAQATWVQGVYLLKTLPLPLPEYVIGLVN 839
Db 826 TGNVEEYSSSAKEILGYTLTTEMPSNAQGEFSLEEQT-VTVIY-SKNPVPKADITV---Q 880
QY 840 YPDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDNEDNEGTHALAVATLADYEGLDIKTIL 899
Db 881 YTDEDG---MELAPTETLSG-----NVDEN-----YVTTAKTFTGYELIETPSNA 922

Db 148 KAEITKQRTAAQENKPNVDLAAIPNVKQIDGKYVYIGSDGQPKKNFALTNNKVLVFDK 207
Qy 216 NEGEVFN--YLLKKVTIPTGKHIGQDAFVNDKNAIEVNLPESETISDYAFH----- 268
Db 208 NTGALTDTSQYQFKQ-----GLTKLNDVTPHQNIVNFEN--TSLETIDNVYTDASWYRP 260
Qy 269 -----LALKQIDL-----PNNLKAIGELAFPDNQITGKLSLPRQLMLAERA 310
Db 261 KDILKNGKTWTASSEDLRPLMSWPDQKQIAYLNYMNQOGLG-----TGENY 310
Qy 311 FKNHIIKTFEPRGSLKLVIGEASPDNDLSQLM-----LPDGLKEIESEAFPTGNP 360
Db 311 TADSSQESLNLAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNNWSQTESDTSAGE- 369
Qy 361 GDDHYNNRVLVTKSGKPSGLATENTYVNDPKSLWQESP--EIDYTKWLEEDFYQKNS 418
Db 370 -KDHLOGGALLYSNDK-----TAVANSYRLLNRTPTSTQTKPKYFEDN----- 413
Qy 419 VTGFSNKGLOKVKRNKNLEIPQHNGVITTEIGDNAFRNVDPNQKTLRKYLDEEVKLPST 478
Db 414 -----SSGGVDFLLN-----DIDNSNPVQAEQLN----- 439
Qy 479 IRKIGAFAPQSNLKSFEASD--DLEETKEGAFNNRIETLELKDCLVTIGDAAAPHNH 535
Db 440 -----WLHYLMNYGSIIVANDPEANPDGVRVDAVDNVNADLLQI-----ASDYLKA 484
Qy 536 IYAVILPESVQBIGRSAPRONGANNLI FMGSKVKTGEMAFPLSNRLEHLDLSEQQLTEI 595
Db 485 HYGVDKSE-----KNAINHLSIL-----EAWSDNDPOY-----NKOTKGAQL 521
Qy 596 PVQAFSDNALKEVLPASLKTIREEAFKQHLKQLEVASALSALSHIAFNALDDNDGD---E 651
Db 522 FI-----DNKRLSLIYALTRPLEKASKN-----EIRSGLEPVITNSLNRSABGKNS 572
Qy 652 OFDNKVVVKTTHNSY--ALADGEHFTVDP--DKLSSTIVDLKILKLEGLDYSTLRQTT 707
Db 573 RMANVIFTRAHDSEVQTVIAKIIQAQINPKTDGLTFLDLEKQAFKIYNEDMRQAKKYT 632
Qy 708 QTOFRDMTAKALLSK-----SNLRQGEKQKFLQEAQFPLGRVLDLDAKAKAKALVTYK 763
Db 633 QS--NIPTAYALMISNKDSITRLYYGD--MYSDDQYMATKSPYDAIDTLKARI--K 685
Qy 764 ATKNGOLLERSI--NKAVLAYNNSAIKANKVRLKELDLTLGLVEGKPLAQATMVQG 820
Db 686 YAAGQDMKITVEGDKSHMDWYTGVLTS-----VRYGTGANEATDQGEAIKTQG 737
Qy 821 VYLLKTPPLPEYIYGLNVYFPGSKLIYALDMSDTIGEGQKDAYGNPILNVDEDEGYH 880
Db 738 MAVITSNPN-----SLKLNQNDKIV-----NMGAAHKQEVYRPLLTTKDGLTSY 783
Qy 881 ALAVATLADY-----EGLDKITLNSKLSQTSIROVPTAAVHRAGIFQAIQNA 929
Db 784 TSDAAKSLYRKNTDKGELVFPASDIQGLNPQVSGYLAV--WVPVGASNDQDVRVAASNK 842
Qy 930 AABARQLLPKPGTHSEKSSSPS-----ANSKDRGLQSNPKTNR 968
Db 843 ANATQVY-----ESSALDSQLIYEGFSNFQDFVTKDSYTNK 881

RESULT 61

B97116
chromosome segregation SMC protein, ATPase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: B97116
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97116
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-1191 <KUR>
A:Cross-references: UNIPROT:Q971A1; GB:AE001437; PIDN:BAK79717.1; PID:g15024720; GSPDB:G
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1751

Query Match 3.5%; Score 175.5; DB 2; Length 1191;
Best Local Similarity 19.2%; Pred. No. 2;
Matches 170; Conservative 154; Mismatches 316; Indels 245; Gaps 42;
Qy 29 KEPI-LKQTASSISGADYAESGSKKLKINETSGPVDVTDLFSDKRTTPEKIKONL 87
Db 262 KEKISLELKVATESLDEPDAKYSSNKTXYESKSEHQILSEIELLEKTSNSDVAKNKL 321
Qy 88 -----AKGPREQELKAVTENTSEKQITSG-SQLEQSKESL-----SLNKTV 128
Db 322 YKEIEDLNSIVNLKRYEIQLTLTEDKNYNKLLSKSEKKKNIDGLIEBEWKS 381
Qy 129 PSTSN--WEICDFITKGNLTGLSGVGEKLSQTDHLVLPQAAADG----- 172
Db 382 KQYKNDIIDIISTISQNNNEVVILKEIE--SNESKLESIKRAGEGYSKSLKINEVTNT 439
Qy 173 --TOLIQVASPAFTPKKTAIAEYTSRAGEN-GEISQLDVPDQKEIINEGEVFNYSLLK- 228
Db 440 LSEELVKI-----NDK--ISGYENQIRENRSKISKLN---RIISDEEKLNNRELNSK 486
Qy 229 -----VTIPTGY-----KHIGQDAFVNDKNIAE-----VNLPESETI 261
Db 487 NKLEANKMLINLEKQYEGYNRSVKNLMQHVTK-GFVDVKPSESVFLGKVIKVKKEFE 545
Qy 262 SDYAFALHALAQIDLPDNL-----KAIGELAFPD-NOITG-KLSLPQRLMLAER 309
Db 546 VEISLGAISDIITLDDNIAKLINLKSQNLGRATFLPLNIIKGRKLNISDATH--EK 603
Qy 310 AFKSNHIKTIIFRGNLSK-----VICEASFDNDLSQLMLPDLKLEKIESEAFTPGPDH 365
Db 604 GFIGIASLIDYDSTFLPAVNVGLRTVI VDNMSALK-----IAKLNS-----Y 648
Qy 366 NRVVLVTKSGKNSGLATENTYVNDPKSLWQESPEIDYTKWLEDFYQKNSVTVGFNK 425
Db 649 SFKIVTLTGEVNVPGSITGSGTSYKAAASITGRKEI-----BELNLELNNVQAAL 701
Qy 426 GLQVKRNKNLEIPQHNGVITTEIGDNAFRNVDPNQKTLRKYLDEEVKLPSTIRKIGAF 485
Db 702 SSNKIENKVI-----VKEL-DMLCLDL---TDTIHGEKIELTKIKERLSID-- 745
Qy 486 AFQSNNL-KSFEAS--DLEETKEGAFNNRIETLELKDCLVTIGDAAAPHNHIAVILPE 543
Db 746 -IESEKLNKSYNTSVGEIGFIKEK--INKHLEKLVKE----- 780
Qy 544 SVQEIGRSAFRONGANNLI FMGSKVKTGEMAFPLSNRLEHLDLSEQKQLTEIPVQA 601
Db 781 --EENKALKLREANNNLI-----DELERKLDKNSKVLNLEINSMK 822
Qy 602 -DNALKEVLLPASLKTIREEAFKQHLKQLEVASALSALSHIAFNALDDNDGDQFONKVV 660
Db 823 VDKAKSDEMLMSSTREIRYKVEHNNMENKISUK-----NEIDD-----FENQVKIL 870
Qy 661 THNSYALADGEHFTVDPDKLSSTIVDLKILKLEGLDYSTLRQTTOTQFRDMTTAGKA 720
Db 871 KNKLEYN-----NGKVQIKQVIELE-----NSFKDSEVERSK 904
Qy 721 LLSKSNLRQGEKQKFLQEAQFPLGRVLDK-----ATAKAKALVTYKATKNGQLLERSI 775
Db 905 L--KGNIE--NKRNSLOGINLVLOKLETEKHRYEINLAKIE-----TESENLYERLN 952
Qy 776 NKAVLAYNNSAIKANKVRL---EKEJDLTLGLVEGKPLAQATM 817
Db 953 SDFKTLVSEAAAEFKEIEBEMISYKKKIDELKREINKMGVNVVASI 997

Db 470 ----NINQOIFKTEVLPFFE-----KGKEQAKLDYGNILNPNYNTQAKVEALPK 517
QY 358 GNPQGDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417
Db 518 GNKNQEIY-----QALDGNVAYEPG 537
QY 418 SVTGSNKGLOKVRKNKNLEIPKQHNQVVTITIGDNAPRNVDFONKTLKRYDLEEVKLP 477
Db 538 AFKSVLNSWTGKIQHPEKADIQRFTHLEQVQIGNSVLN-----OPQ 580
QY 478 TIRKIGAFQFQSNLKSPEASDDLEBEIKGAFMNNRIETLELKDKLVTIGDAAFHINHIY 537
Db 581 ITKEQVSSLSKNF-----FKNGHQVASYFQDLLTKDKL-TVLETLYDLAKKW 628
QY 538 AI-----VLPEVQEIGRSAFRQNGANNLIFMGSVKVTIGEMAFLSNRLHLDLSEOKOL 592
Db 629 GLETNWAQPKAFQYTKDIFAB--ADKLKFLWKK-----DPYNOI 669
QY 593 TEIPVQAF-----SDNALKEVLLPASLKTIRREAFKQNHKQLEVASALSHPN 642
Db 670 NEIHQLSFNILARNVDIKSDGFYGVLLLPQSVKT---ELEGKVEAQ-----IFE 715
QY 643 ALDDNDGDEQFQNVVVKTHNSYALADGEHEFIVDPDKLSSTIVLEKILKILIEGLDYST 702
Db 716 AL-----KKYSLIENSAP-----KTTILD-----KNLLEGTDFKT 745
QY 703 LRQTTQTOPRDMTTAGKALLSKSNLRQGEKQKFLQEAQF--FLGRVDLDKAKAEKALV 760
Db 746 -----FGDPL-----KAFFLKAAQNNFAPWAKLDDNLQYFEAIK 781
QY 761 TKKATNGQLLRSINKAVLAYNNSAIKKANYKRLKELD-LLTGLV-----EGKGP 811
Db 782 KGBTKEG-----KREEDVKVKELDNKIKGILPOPPAAKPEAAKP 822
QY 812 LAQATWQGVYLLKTPPLPEYIYGLNVYFDSKGLIYAL-DMSTIG--EGQKDAYGN 867
Db 823 VAAKPVAAPKPEAAKPVAAPKPE-----AAKPVAAPKPEAAKPVAAPKPEAAKP 871
QY 868 PTLN-----VDEDNEGYHALAVATLADY---EGLDKT-----TLNSKL 903
Db 872 PVATNTNTNTGFSLTNKKPKEDYFPMAFSYKLEVTDENKLSLKTPEINVFLELVHQSEY 931
QY 904 SOLTSIRQVPTAAYHRAGIFQIAONAAEAEOLLPKPGTHSSKSSSANSKDRGLQSN 963
Db 932 EDOKIIEKDKTVLNLQYQFQVVKVTSQYQKLSHPMTEGTQNGQKKGEGTPNQKKAE 991
QY 964 PKTNRGRHSAILPRTGSK 981
Db 992 GAPSQKKAEGTPNQKK 1009

RESULT 64
AG1717
probable peptidoglycan bound protein (LPXTG motif) lin2283 [imported] - Listeria innocua
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AG1717
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AG1717
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1151 <GLA>
A:Cross-references: UNIPROT:Q929J2; GB:AL592022; PIDN:CAC97511.1; PID:g16414795; GSPDB:C
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin2283

Query Match 3.4%; Score 175; DB 2; Length 1151;
Best Local Similarity 18.9%; Pred No. 2;
Matches 234; Conservative 171; Mismatches 408; Indels 424; Gaps 54;

QY 8 VALTITVSVVTHNQE-VFSLVKEPILKQTOASSISGA--DYAESGSKSLKLINE-TSG 63
Db 102 IYLIKTPNTIYKNAKLVINLPPKNGEFTQSLSELKAGVTPSYNKTTRQLRVVLYKYOTLNSG 161
QY 64 PVDDTVTDLFSDKRTTPEKIK-----DNLAGPREQELKAVTENTESEKQITSGSQ 114
Db 162 VVDKVLKSLKSTNGYTPNGTKLEVTGBFSADNL-----EKVTEQAETTVNATATTA 213
QY 115 L--EOSKESLSLNTKVPSTSNWEICDFTTKGNTLVGLSGSGVEKLSQTDHVLVPSQAADG 172
Db 214 LSNEFTKVENSLNN--PSQGDIGISF-----NLNIPKSTGS 250
QY 173 TQLIQVAFAPTPDKKTAIAEY-----SRAGENGEISQLDVDGKI-----IN 216
Db 251 -----LFTIEGKIIIEYTLADKLDYLGAGDTPEPTK--TEGQKLTWEIAAPTLY 299
QY 217 EGEVFNYSLLKKVVTIPTGYK-HIGQDAFVDKNKNAEVLNLPESLETSIDYAFALHALKQID 275
Db 300 BOEKATSLNKTFFQIRTFQTSIPNFATVENKAVATTNFTLTDSDIVDTSKASVSASD 359
QY 276 LPDNLKALGELAFDQITGKLSPLQRLMLRAERAFKSNHIKTIIEFRGNSLKVIGESAPQ 335
Db 360 PATIPPTIGSV-----YAPAHRCPVDPENWGIAVTGPNDIK 395
QY 336 DNDLSQMLPDLGLEKIESEAFTPNGDD--HYNNRVLMWTKSGKNPSGLATENTYVNPDK 393
Db 396 AYDSAKL-----GFSLMLNSAMNDSPSYFIYV-----DVYVNIID 431
QY 394 SL-----WQESPEIDYTKWL-----BEDFTYQKNSVTGFSNKGLO 428
Db 432 NLNLDYFRSGDFYFKPNVNYPCWQPKKSPKYNLLVKYDGDNDMTLTKEDVE-----LS 485
QY 429 KVKRNKNLEIPKQHNQVITBI-----GDNAFRNVDFQNTLRYKDYLEEVKL 475
Db 486 KMYSEKDLGIDDDKH---VSKWHLFTYAPAGMYSADLSFFT-----TVKEGYGEVRN 536
QY 476 PSTIRKIGAFQSNLKSFEASDDLEE-----IKEGAFMNNRIETLE 518
Db 537 TTKINMYGADS--QGYIHYFDDTNPWPENWQKYAGDRTAQIIPOPIGKNKFKVQGSVAPDD 594
QY 519 LKDKLVITGDAAFHINHIYAIVLPESVQEIERSAFRQNG---ANNLI FMGSKVKTLGEMA 575
Db 595 TDGNLINIGDNSVSVN-----LESNKASISRLTGPPEAMVLLPSPGVMKNTDQOG 644
QY 576 FLSNRLHLDLSEOKQLTEIPVQAFSDNALKEVLLPASLKTIR----- 618
Db 645 FNVSVLNSNYQNSGRQLLKVKWD-----KKTLLPAEKLTAKNVTVSKDTPSNMVBEM 697
QY 619 -----EEA-----FKQNHKL-- 628
Db 698 FGFLQDQDFNVPEVTGPTISDTKWEIDSNINQNGNSEESRITSGNHVILNTSNHLKIS 757
QY 629 -----OLEVASALSHIAPNALDDNDGDEQFQNVKVVKTHNSYALADGEHIV 676
Db 758 KKAAGNLDKGYSGLANATTNSIVSYQLLENDSDEKIANMVMMDVLPSEKDLG-----IT 812
QY 677 DPDKLSSTI-VDLKILKL-----TEGLDYSTLRQTTQTFRD-----MTTAGKALLSK 724
Db 813 DNSRGSKFNLELTKAVAI PNWDKVB-VVYSTAKNPKRSGVLVDKHTIYPTGSPFLVDN 871
QY 725 SN-----LRQGEKQKFLQEAQFFLGRVDDLKAI-----AKAEKALVTTKATNGQLLRS 774
Db 872 REATEADWLTAQSEVDWSEKHSF---XWELKEGIEWIPGKSMKIQFDLTKPKQNDKEL 928
QY 775 INK-----AVLAYNNSAIKKAN-----VKRLSEKL---DLL 802
Db 929 LNQQTKKEDRAAWNSFAVAVNNNSQVIEPAQGVVALDSDSVA PVTYQVYDQNHKQIASPETL 988

```
Qy 803 TGLVEGKGLAQATWVQGVYLLKTPPLPEYVIGLVVYDPKSGKLIYALDMSDTIGEG-- 860
Db 989 TG-AYGEFTAKQKITWYLVKTP-----ANVSGTFNEKAQ 1024
Qy 861 -----QKDAYGNPILN-VBDNEGYHALAVATLADYEGLDIKTILNKL--SOLTSIRQ 911
Db 1025 TVTFVYQKVTAGNIIVDYDNKE-----KLAD-----SIVLTGKLSVYTSACK 1070
Qy 912 VPTAAHRAQIFQAIQNAABAEQLLPKPGTHSEKSSSESANSKDRGLQGNPKTNRRGH 971
Db 1071 I--SGYK---LYQTPKNAAGKFSNTSRVTYVYKASTF-----SISNKGADNVEK 1117
Qy 972 SAILPRTG-SKGSFVYGLGYTSVALLSLITAIKKK 1007
Db 1118 SAKUPQTGDSKKTNLWPIFG---MLLSSTTFVIMKK 1151

RESULT 65
A43336
microtubule-vesicle linker CLIP-170 - human
C/Species: Homo sapiens (man)
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
C/Accession: A43336
R/Pierre, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.
Cell 70, 887-900, 1992
A/Title: CLIP-170 links endocytic vesicles to microtubules.
A/Reference number: A43336; PMID:92405160; PMID:1356075
A/Accession: A43336
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1392 <PIE>
A/Cross-references: GB:M97501; NID:g180621; PIDN:AAA35693.1; PID:g180622

Query Match 3.4%; Score 175; DB 2; Length 1392;
Best Local Similarity 19.6%; Pred. No. 2.7;
Matches 191; Conservative 177; Mismatches 391; Indels 216; Gaps 40;

Qy 2 KKHLLKTVALTITVSVVTHNQEVSLVKEPILKQTOASSISGADYAESSKSKLKINET 61
Db 435 KRKVEDLQFRVEESITKGDLEAVATVSEKSRIMELEKDLALRVQEAEL--RRRLSNKP 492
Qy 62 SGPVDVTDVDFSKRTTPKIKONLAKGPREOLKAVTEN----TESEKQI---TSG 112
Db 493 AGDVMSLS--LQELSSIQLEKLE--VTRTDHQREITSKEHFGAREBTHQEKIYATAT 549
Qy 113 SLEQSKESLSLNTKVPSTNWEICDFI----TKGNTLVGLSKSGVEKLSQTDHLVPSQ 168
Db 550 EKLKSENESS--KSKLEHANKENSVDIALMWSKLETAIASHQQAWEELKVSFSGLOT 606
Qy 169 AADGTQLIQVASFAPTPDKTAIABYTSRAGENGEI-----SQLDVGDKEIINEGEVFN 224
Db 607 TAEFAEL-----KTQIEK--MRLDYQHEIENLQOQDSERAAHAKEMEALRAK 652
Qy 225 LLKKTPTPTGKHIGQDAPVDNKNIAE-VNLPESLETISDYAFALHAKLOIDL----- 276
Db 653 LMKVIK-----EKENSELAIRSKLDKAEQHLVEMEDTLNKLQEAIEKVELEVLOAKNE 708
Qy 277 -----PDNLKATGE-----LAFPDNQITGKLSLPRQMLRAERAPKSNHIK----- 317
Db 709 QTKVIDNFTSOLKATEEKLLDALARKASSEGSEMKKLRLQLEAAEQIKHLEIKNAE 768
Qy 318 -----TIEPRGSLKVIQGEASFOQDNLSQLMPLDGLEK-----TESEAFNGPDHNN 367
Db 769 SSKASSITRELGRELKLE---TNLQENISEVSVQKTELEKELQILKEF----- 814
Qy 368 RVLWTKSGKPNGLATENTVNPDKLQWSPREIDYTKWLEEDFTYQKNSVTGSKGL 427
Db 815 -----REASEEAVSVORSQETVYNKLHQEEQFN---MLSSDLEKLENLADWEAKFR 864
Qy 428 QKVKKNKLEIPKQHNGVITFEI-----GDNAFRNVDFQNK-TLRKYDLEEVKLPSTIRKI 482
Db 865 EKDEREEQLIKAKELENDIAEIMKMSGDSSQLTKMDELRLKRDVEELQLKLT----- 920
```

```
Qy 483 GAFAPQSNLKSF--EASDDL-----EEIKEGAFNNRIETLEKDKLVITGDAAPHINHI 536
Db 921 -----KANENASFLOKSIEMDTVKAEOQOBAKKEHEEKELERKLSDL----- 965
Qy 537 YAIVLPESVOBGRSAFRQGANNNLIPMGSKVKTGLGEMAFLSNRLHLDLSSEQQLTEIP 596
Db 966 -----EKKMETHNQOE-----LKARYERATSETYTKHEELILONLQKTLDD 1007
Qy 597 VQAFSDNALKEVLPAASLKTIREAFKQKNHLKQLEVASALSHTAFNALDNDGDQPDNK 656
Db 1008 TE-----DKLKGAEE--NSGLLOLEELRQKAEKAKAAQTAEDAMQIMEQM 1052
Qy 657 VVKTHNSYALADGEHFIVDPDKLSSTIVDL-EKILKLEGLDYS-----TLRQTTQTOF 711
Db 1053 TKEKTE-----TLASLEDTKQTNAKLQNELDTLKENNLKNVEELNKSKELLTVENQKMEEF 1108
Qy 712 R-DMTTAGKALLSKS---NLRQGEKQKQFLOBAQPFQGRVDLDKAIKAEKALVTKKATKN 767
Db 1109 RKEIETLKQAAQKSQQLSALOENVKLAE---LGR---SRDEVTSHQKLEBESVLN 1161
Qy 768 GQLLE-----RSINKAVLAYNNSAI---KQANVKRLEKELDLTLGLVEG 808
Db 1162 NQLEMKKRESKFKTDADEEKASLQKSI--TSALLTEKDAELEKLRNEVTVLRG--EN 1217
Qy 809 KGPLAQATWQGVYLLKTPPLPEYVIGLVVYDPK-----SGKLIYALDMSDTIGEGOK 862
Db 1218 ASAKSLHSVVQTLSESDVKVLELQVKNLELQLENKROLSSSSGNTDQOADEDERAQSQI 1277
Qy 863 DAYGNPILNVDENE 877
Db 1278 DFLNSVIVDLQKNQ 1292

RESULT 66
T01553
hypothetical protein A_TW018A10.17 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 09-Jul-2004
R/Accession: T01553
R/Dempsey, S.; Harper, M.
submitted to the EMBL Data Library, July 1997
A/Description: The sequence of A. thaliana TW018A10.
A/Reference number: Z14348
A/Accession: T01553
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-976 <DB>
A/Cross-references: UNIPROT:O23084; EMBL:AF013294; NID:g2252848; PID:g2252870
A/Experimental source: cultivar Columbia
C/Genetics:
A/Map position: 4
A/Introns: 116/1; 200/1; 344/1; 447/1; 529/3
A/Note: A_TW018A10.17

Query Match 3.4%; Score 174.5; DB 2; Length 976;
Best Local Similarity 20.1%; Pred. No. 1.6;
Matches 141; Conservative 114; Mismatches 255; Indels 193; Gaps 31;

Qy 6 KTVALTITTV-SVVTHNQEVSLVKEP-----ILKQTOASSIS-----G 44
Db 343 KETELGVKTVEDVGTNNMETDVLVSHPNHLVANSELSTPVTKEKKRQKSKDQINLS 402
Qy 45 ADYAESSGKSLKINETSGP-----VDDTVTDLFSDKRTT-----PEKIKONLAKGP 91
Db 403 ATAATATTASREIVNEINGAPGNVDVDATESECLISQRENLGKENGEBIKEDVSGG- 461
Qy 92 REQELKAVTENTSEKQI-----TSGSQLEQSKESL-----SLNKTVPSTN 133
Db 462 --KAIPSAADNIQTDKKVSYPDNQLEVSGLTGTTTEKKRQKSKKNIDNSTAATTT 519
Qy 134 WEIC--DFITKGNLTVGLSKSGVEKLSQTDHLVPSQAADGTQLIQVAS--FAFTPPDKT 189
Db 520 SKIIVNERKMGDILVDKVGSGREATLIGADKI-----QAASDLHVAGMASTPHAFVQESKT 575
```

QY 190 A--IAEYTSRAGE--NGEISQDVGKEIIINEGEVNSYLLK-----KVITPTGYKHI--- 238
Db 576 LDHIGKVTDESEHPKVERVAIDADQAKSVKTSKSRKAKTPAKEDTLVDGSAQNVPEPI 635
QY 239 -----GQDAFVDN-----KNIAEVNLPESLETISDYAFALHAKQIDLPDNLKAIGE 285
Db 636 KVVDCGEGHDNVRINVLDSIQORNEAENMEKSGKSS-----KRSKKDSLNIIVEE 686
QY 286 LAFFNOITGKLSLPRQLMRLAERAFKSNHIIKTIFRGNLSKVIGEA---SFQDNDLSQL 342
Db 687 AQVUS--LOOKEAENLEKSGKSKRSKK-----DSLNIIVEAQVLSVEVNVNQ- 738
QY 343 MLPDGLKEKTESAFTGNPDHNNRVNLTGSKNPSGLA-----TENTY-VNPD 392
Db 739 -----EASPINPKTD-----ALFPAKKNSTESNPLKKIIEVTINTEDINRS 784
QY 393 KSLWQSPRIDYTKMLEEDFTYQKNSVTGFSNK-----GLQKVRKNKNLEIPKOHNGVT 446
Db 785 MQVQKENAGMG-----DNIGSQKDDIVGANKQDQVGTGTSKKEKSLDL---HPGGS 836
QY 447 ITSIGNAPRNVDQFNKTLKYDLEEVKLPSTIRKIGAPAFQSNLKSFEASD-----D 500
Db 837 ID-----GSMKMKETKGRVQPSSTGTSQLOSNMKNDRSGSKVD 874
QY 501 LEEIKEGAFPMNRIETLELKDKLVTIGDAAFPHNIIYAIVLPSVQEIGRSAFRQNGANN 560
Db 875 LSDAPMKGTNNKKAIVKSSKSVTVNKSMMVNN-----KEKAVKISNVTANKSTTN 929
QY 561 LIFMGSKVKTGEMAPLNSRLHLDLSEQKQTEIPVQAFSDN 603
Db 930 -----PFKDAEBDESKTTSSTDSTKAPSDSSDN 957

RESULT 67
T13030
microtubule binding protein D-Clp1-190 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13030
R:Lantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A:Title: A class VI unconventional myosin is associated with a homologue of a microtubul
A:Reference number: Z17588; MUID:98139549; PMID:9472041
A:Accession: T13030
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1690 <LAN>
A:Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AAB96783.1
A:Experimental source: strain Oregon R
C:Genetics:
A:Cross-references: FlyBase:FBgn0020503
C:Keywords: cytoskeleton

Query Match 3.4%; Score 174.5; DB 2; Length 1690;
Best Local Similarity 19.7%; Pred. No. 3.7;
Matches 244; Conservative 189; Mismatches 448; Indels 359; Gaps 54;

QY 3 KHLKTVALTTLTVSVVTHN-----QEVSLVKPEILKQTOASSISGADYAESGK 53
Db 460 KITLVSATPSLQSIPLPDLPSDDGALQBEIAQLQEKMTIQQKEVESRIAEQLEBEQRLR 519
QY 54 SKLK-INETSFPVD-----DVTVDLPSDKRTTPEKIKDNLAKGPREQELKAVTENTSE 106
Db 520 ENVKYLNQIATLQSELVSKDALBKFSLSECGIENLRELALLKEENEKQAEQAEFT 579
QY 107 KQITSGS--QLPQSKESLSNKTVPSTSNWEI-----CDFITKNTLVGLSKSGYEKLS- 158
Db 580 RKLAEKSVLRLSELQNLKATSDLSLSESRVKNKSDECEIL---QTEVRMRDEQIRELNQ 636
QY 159 QTDHLV--LPSQAADGTOLIQVAFPT--PDKTAIAEYTSRAGENGISQLDVGKBI 215
Db 637 QLDVTTQLNVQKADSSALDMLRLQKEGTEKSTLLEKTEK-----ELVQIKEQAANTL 691

RESULT 68
B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)

QY 216 NEGEVNSYL--LKKVTTIPTGYKHIGQDAFYDNKN---IAEVNL--PESLETISDYAFALH 269
Db 692 QDKEQLEKQISDLKQLA-----EQKLVREKTENAINQIQLEKESIE-----QOL 736
QY 270 ALKQIDLPDNLKAIGELAFFDNQITGKLSLPRQLMRLAERAFKSNHIIKTIEF--RGNLSK 327
Db 737 ALKQNELEDPOKQKQSESEVHLQEIKAQ-----NTQKDELVESGESLK 779
QY 328 VIGEASPDNDLSQLMLPDGLKEKTESAFTGNPDHNNRVN-----LWTKSGKN 378
Db 780 KL--QQOLEEKTGLGHEKLALEELKEKET-----IIEKEQEOLOQLOKSAS 827
QY 379 PSGLATENTYNNP-----DKSLWQSPRIDYTK--WLEBDFYQKNSVTGFSNK 425
Db 828 ESALKVVOVQLEQLOQQAASGEGSKTVAKLHDEISQLKSAETQSELKSTESNLEAK 887
QY 426 GLQKVRKNKNLEIPKOHNGVTITIGDNAPRNVDQFNKTLKYDLEEVK--LPSTIRKIG 483
Db 888 SKOLEAANGSLSEBAKSGQLQEQI-----TLKSEVEETOAAALSSVHTDV- 933
QY 484 AFAPQSNLKSFEASDLDLEEKEGAFPMNRIETLEKDKLVTIGDAAPHNIIYAIVLPE 543
Db 934 -----ESKTKQLEAANAALAEKVNKEYAESRAEASDLQDKVKEITDT-----LHAELOAE 982
QY 544 ---SVQEIGRSAFRQNGANNLIIFMGSKVKTGEMAPLNSRLHLDLSEQKQTEIPVQ- 598
Db 983 RSSSALHTKLSKP-----SDEIATGHKELTSKADAWSQEMLQ-----KEELQELRQOL 1032
QY 599 -----AFSDNALKEVLLPASLKTITREBAFK-----KNHLKQLEVA 633
Db 1033 QDSQDSQTKLKAEGERKEKSPESIKNLQEEVTKRAKTENLELSTGTQTTIKDQERLEIT 1092
QY 634 SA--LSHTAFNALDNDGDEQFDNKVV-----KTHH 663
Db 1093 NAELOH--KEKVASEDAQKIADLKTLEBAIQVANANISATNAELSTVLVLEQLAKSETNH 1150
QY 664 -----NSVALADGEHFIVDPDKLSSTIVDLKILKIEGLD-----YSTL 703
Db 1151 IFELFEMEDAMNSELRIEKVTGI--KEELKETHIQLDERQKRFEELEKILKQAQOSEKL 1208
QY 704 RQTTQTOFRDMMTAGKAL-----LKSNNLRQ 729
Db 1209 QOESQTSKEKLEITEIQOSIQELQDSVKQKEELVQNLEEKVRESSSIIIEAQTKNLSENVQL 1268
QY 730 GEKQKFOEAOFFLGRVDLDKAI--AKAEKALVTYKATYNGOLLERSINKAVLAYNNSAI 787
Db 1269 ENKTSCLKETQ-----DQLLESQKKEKQLOBEAAKLSGEL--QQVQEAANGDIKDSL 1318
QY 788 K-KANVKELEKELDLTLGLVEGKGLAQAT-----MWQGVYL----- 823
Db 1319 KVESLVKVELEKQAATSQLDQAQ-----QATNKEIQELLVKSQENEGNLQGESLAVTEKLQ 1374
QY 824 -----LKTPLPLPEYIIGLVNVPFKSGKLIYALDMSDTIGECQKDAYGNPILNVDED 875
Db 1375 QLEQANGELKEALCQKE-----NGLKELOK-----LDESNTVLESQKSHNEIQDKLSQA 1425
QY 876 NEGTHALAVAT--IAD-----YEGLDITKLINSKLSQITS-----IROVP 913
Db 1426 QOKERTLQEETSCLAQESQLKQANEELQKSLQOKLLEKGNBFTQLABYQKVQIDEMD 1485
QY 914 TAAVHRAGIFQAIQNAAEAEQLPKPCTHSEKSSSESAN-----SKORGLOSN 963
Db 1486 DAASVAGLLEQLQNRVAELETAL--RQANDAQKTAYLETKEILRQLESELEKREVLUSL 1544
QY 964 PKTVRGHSAILPRTGSGSFVYIGILYTSVALLSLITAI 1003
Db 1545 KAQWNGASS-----RSG-KGDEVESLDIETSLAKINFLNSI 1579

Db 585 LKEDYVPSLEYLVEHAKATNHHLLSDSAYEDLVKCKENPDMEFLKESAKLGHVTVSN 644
QY 386 NTVNPKSLWOBSPIDYTKWLEEDFTYQKSVTFSGNKGLOKVRKNLEIPKOHNGV 445
Db 645 EAYSELEKKL--EOPSELYL--VEHAKATNHHLLSDSAYEDLVKCKENPDMEFLKES-- 698
QY 446 TITEIGDANFRNVDFONKTLRY-DLE-EVKLPSTIRKIGAFQFQNNILKSFASDDEE 503
Db 699 --AKLGHVTVSN-EAYSELQRYSELEKEVEQPSLAYLVEHAKATDHHLLSDSAYEDLVK 755
QY 504 IKEGAPMNRRIETLEKDKLVITGDAAFHINHIYAVLPESVQETORSAPRONGANNLIF 563
Db 756 CKE---NPDVEF--LKEKSAKLG-----HTVVSSEYSELQRY----- 789
QY 564 MGSKVKTGEMAFLSNRLEHLLDSEKQLTEIPVQAFSDNALKEVLL---PASLKTIREE 620
Db 790 --SELEKEVEQPSLAYLVEHAKATDHHLL-----SDSAYEELVKCKENPDMEFLK 839
QY 621 AFKONH-----LKQ-----LEVASALSH--IAPNALDD-----NDGEQF 653
Db 840 SAKLGHVTVSNSEAYSELEKKLEQPSLAYLVEHAKATDHHLLSDSAYEDLVKCKENSDVEF 899
QY 654 DNKVVVKTNN-----SYALADGEHFIVD-----P 678
Db 900 LKEKSAKLGHTVVSNEAYSELEKKLEQPSLAYLVEHAKATDHHLLSDSAYEDLVKCKENP 959
QY 679 D-----KLSSTIVDLKILKLEGLDYSLRTTOTQPRDMTTAGKALLSKS---- 725
Db 960 DMEFLKESAKLGHVTVSNSEAYSELEKKLEQPSLEYLVE--HAKATNHHLLSDSAYED 1015
QY 726 --NLROGEKQFLQEAQFFLGR-VLDLKAIAKAELAVT-----KKATNGQLLER 773
Db 1016 LVKCKENPDMEFLKESAKLGHVTVSNSEAYSELEKKLEQPSLEYLVEHAKATNHHLLSDS 1075
QY 774 STNKAVLYNN---SAIKANVK-----RLEKELD--LITGLVE-----GKQP 811
Db 1076 AYSELVKCKENPDVEFLKESAKLGHVTVSNSEAYSELEKKLEQPSLEYLVEHAKATNHHLL 1135
QY 812 LAQATWVQGVLLKTPLEPYIGLVYF--DKSGKLIYALDMSDTIGEGOKDAYGNPI 869
Db 1136 LSDSAYEELVKCKENPD-----DVEFLKESAKLGHVTVSNSEAYSELEKKLEQPSL 1185
QY 870 LNVEDNEG--YHALAVATLADY-----EGLDIK-----TILNLSKLSQTSIRQ 911
Db 1186 AYIVEHAKATDHHLLSDSAYEDLVKCKENPDVEFLKESAKLGHVTVSNSEAYSELEKKLE 1245
QY 912 VPTAYHRAGIFQATQNAAEQQLL 937
Db 1246 QPSLAY-----LVEHAKATDHHLL 1264

RESULT 73
T14156
Kinesin-related protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T14156
R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.
Cell 91, 357-366, 1997
A:Title: KENP-E is a plus end-directed kinetochore motor required for metaphase chromosome segregation
A:Reference number: Z17893; MUID:98028574; PMID:9363944
A:Accession: T14156
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2954 <WO>
A:Cross-references: UNIPROT:O42263; EMBL:AF027728; NID:G2586071; PIDN:AA68
C:Genetics:
A:Gene: KENP-E
C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 3.4%; Score 173.5; DB 2; Length 2954;
Best Local Similarity 20.1%; Pred. No. 9.5;
Matches 209; Conservative 186; Mismatches 431; Indels 213; Gaps 47;

QY 23 EVFSLVKEPILKQTOASSISGADYAESGSKLAKINETSQPVDDTVTDLPFSDKRTTPEK 82
Db 1200 EVKLELLEWELKHGHTDSQLS-----IEKQLENLEVTETKLTQLOBEMKNITERNELQTN 1255
QY 83 IKDNLAKPREOELKAVTENTESEKQITSGSQLEQSKESLSLNKTPVSPSTNNWEICDFITK 142
Db 1256 PED--LKAHDSLKQDLSENI--EQSIETQDELRAAQEELREKQQLVDSFRQQLLD---- 1307
QY 143 GNTLVGLSK-----SGVEKLSQTHLVLPQAAAGTQ-LIQVASFAPTPDKKTAIAEYT 195
Db 1308 --CSVGISPNHDAVANQEKVSLGEVNSLOSEMRLGRDELQTSCKALVSELELRAHVK 1365
QY 196 SRAGNNGIS--QLDVGKEIINEGEVFNFSYLLKVTIPTGYKHIQODAFVDMKNKIAEVN- 253
Db 1366 SVEGENLEITKNGLEKEILKSE--ESEVLKSM-----ENKEDNN 1407
QY 254 -LPSELETISD-----YAFALHALKQID-LPDNLKA-----IGELAFFD----- 290
Db 1408 KLKEQAEVSSKENQFSLSEEVFSGQLVDEIVLKAQLKAAEERLEIKDRDYFELVOTA 1467
QY 291 --NQITGKLSLPROLMRLAERAFKSNHIK-TIEPRGN--SLKVIGEASFQNDL----- 339
Db 1468 NTNLVEGKLETPQ-----ADHEEDSIDRSEEMEIKVLGKELRNQYLLERLQ 1517
QY 340 SOLMLPDGLEKTESAFTGNPGDDHNNRVVLTWTKSGKNPSPGLATENTVV--NPDKSLWQ 397
Db 1518 EKLELSNKLEILOKEMETS-----VLLKDDLOQLKLSLSENILKENIDITLKH 1567
QY 398 ESPEIDYTKWLEEDFTYQKSVTFSGNKG--LQVKRNKNLEIPKOHNGVITTEIGDNP 455
Db 1568 HSDTOAQLOKTOQLAKLAIASDNCPIQEBKETSADCVHPLEEKILLITE---ELH 1624
QY 456 RNVDPOKTL--RYDLE--EVKLPSTIRKIGAFQFQ--NNLKSPF--ASDDLEIEKEGAPM 510
Db 1625 OKTNEQKLEHKEKLEQAQVELKCEVHLMKSMIESKSSLESLOHEKHDTQQLLALQ 1684
QY 511 NNRIETLEKDKLVITGDAAFHINHIYAVLPESVQETIGRSAFRONGANNLIFMGSVKVT 570
Db 1685 QMNVVTKKLEQLQTHLTAEDVH-----LKENI--EUGLN--FKNE-----AQQT 1728
QY 571 LGEMAFLSNRLEHLLDSEKQLTEIP--VOAFSDNALKEVLLPASLKTIREEAFKKNHLK 628
Db 1729 TKEOCLL--NENKLEQSQHRIQCEIEELMKSLKQ--KE---SALETKSEQKVINLN 1780
QY 629 Q-----LEVASALSHIAFNALDDNDGDSQFQFNKVVVKTNN-----SYALADGEHFIV 676
Db 1781 QEMEMVLEMEELKNSQRTVIAERDQLQDLRESVEMSIETQDDLKQAEALQOQKQKV- 1839
QY 677 DPKLSSTIVDLKILKLEGLDYSLRTTOTQPRDMTTAGKALLSKSNLRQEKQKF- 735
Db 1840 --QELTSQISVLQEKISLGE-----NQMLYNVATVKETLSEKDDLNQSKQLHFS 1886
QY 736 -----LOBAQFFLGRVLDLKAIAKAELVTKKATK--NGQLLERSINKAVLAYNNS- 785
Db 1887 EITLSLSLKEKEFALEAKDAKADAKRTIDITEKISNIEBQLLQOATNLKETLYERES 1946
QY 786 -----AIKKANVRLEKELDLTLGLVEKGPQAQATWVGQVYLLKTPPLPYEYGL 837
Db 1947 LIQCKEQLALNTEHLRETLKSDALGKMEQERDEA---ANKVIALTEKMSLSLEEQINE 2002
QY 838 NVYFDKSG---KLIIY-----ALDMSDTIGEGOKDAYGNPILNVDS-----DNEGYHA 881
Db 2003 NVTTLKEGEGEKETFLQRPQSKSSSQMEELRESLKTKDLOLEBAEKEISEATNEIKNL 2062
QY 882 LAVATLADYEGIDIKTILNLSKLSQTSIRQVPTAYHRAGIFQATQNAAEQQLLKP 941
Db 2063 TAKLSLEEEILQNASILNEAVSERENLR-----HSKQQLVSELEQL--SLT 2107
QY 942 THSEKSSSESSEANSKDRGL 960
Db 2108 LKSRDHAPAQSKRENDEAV 2126

```
RESULT 74
T30996
Hypothesis: protein C01G5.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30996
R:Bradshaw, H.; Stellyes, L.
Submitted to the EMBL Data Library, August 1999
A:Description: The sequence of C. elegans cosmid C01G5.
A:Reference number: Z20956
A:Accession: T30996
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1079 <BRA>
A:Cross-references: UNIPROT:Q17566; EMBL:U50068; PIDN:AAB37736.2
A:Experimental source: strain Bristol N2
C:Genetics:
A:Map position: IV
A:Introns: 10/1; 31/1; 75/1; 108/1; 144/1; 927/2; 1056/2
A:Note: C01G5.4

Query Match          3.4%; Score 173; DB 2; Length 1079;
Best Local Similarity 20.3%; Pred No. 2,2;
Matches 202; Conservative 152; Mismatches 410; Indels 230; Gaps 44;

QY 149 LKSGVEKLSQTDHVLPSQAADQGLIQVASFAPTPDKKTAIABYTSRAGE-----NG 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 VSSGSPDAASQDYQLIGQSSSDHVSLSLATPLSKHAPDRRSFLKRSASGHESPSPHNA 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 203 EIS-----QLDVGGEINEGEVFNFSYLLKKVTITGY-KHIGQDAFDVN-----KNIAE 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 DVSPSSPOOLHLAFANATSTNTVNA---KSTNFDGYLKHISVIARHANGITLQAGLMN 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 252 VNLP-----ESLETISDYAF-----AHLAKQIDLPDNLKALGELAFDNPQITG 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 EKIPIDAVVGELFGLRSIE-LSDIYFKLDKIAWTKLKDIPNDLKASTDELDELESKI-- 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 296 KLSLPRQLM-----RLAERAPKSNHIKTIIEPRGNSLKVIGASQD--NDLSQLMLPDG 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 -LELDQIRMESDGVQNVLESKLTDFATYKVNKNKYNKPLPSDANREISTVLIT-- 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 348 LEKIESEAFNGPGDDHNNRVVLTAKGKPSGLATENTYVNPDKSLWQSPEDY--- 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 LEKILLSATAKE-----LWNKYSRLSPSCCKGMAATEAVKKIDKKXIEKYGEM 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 405 -----TKWLEB--DFTYQKNSVTGFSNKGLOKVKR--NKNLEIPKQHG----- 444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 337 AVGATTFPRLQVIDLWRRER---YSNGKANTNMVNEANLEKYNELNANDASQDFKQI 393
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 445 ---VTITEIGDNAPRNVPQNTLRKYDLEEV-----KLPSTIRKIG--APAFOSNNLK 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 394 LELVNAEQDHSKMTIGFSRGIL--DLKHVAGDARDQWLSNVLNLSDEHPARITSGILQ 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 494 S-FAASDDELEETKEGAFNNRI---ETLEKDKLVITGDAAPHINHIYALVLP-----ES 544
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 PULKASNDLETNDKMLTVRSVALSQSLVALSSTVITTSYSSLSKGFKLINDYDECER 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 545 VQIEIKSAPFRONGANNLIFMGSKVITLGMAPLSNRLEHLDL--SBOKOLTEIPVQAFSDN 603
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 IQSVTTDYIRSGPITLIRPKDQVQKPINEMLEIAMKLDVEGLNAYAKENEL--NKYGDG 568
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 604 ALKEVLLPAS---LKTIRREAFKKNHLKQLEVASALSHAFNALDDNDGDFQDNKVVVK 660
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 569 G-----PHSDQOLETIHENLQOGSLTNLQNRVAGIQTLSGFDVNN---FNSVTLSK 618
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 661 -----THNSVALADGEHFIVDPDKLSSTIVLEKILKLEGLDYSLTRTQTQ 710
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 619 ADSVQAGMDDTRGIMKVADEMNFIQCTQKSSSDSVMLSKGVRAIQ-----RLRNVTADE 673
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 711 FRDMTTAGKALLSKNLRQGEKQKFLQBAQFFLGRVLDLDKATAKAEKALVTYKATNQGL 770
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 674 IEPLET-----STFFSELSFKELAAIKSIPKMKNGASKETLE 711
```

```
QY 771 LERSINK-AVLAYNNSAIKKANVKRLEKELDLLTGLVEGKPLAQATWVG-----VYLL 824
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 712 LNRMAQSLSAIEYSASALKSA-----YGLLELESSIGQLKNVDAAVSGEIRKI 759
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 825 KTLPLPEYVIGLVNVPDKSKGLIYALDMSDTIGEGOKDAYGNPILAVNDEDEGYHALAV 884
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 760 KNPADIKKLQ---NQWGD-----HKDMGSLNANAKAYFYVMINIVSK-----L 801
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 885 ATLADYE-----GLDIKTILN--SKLSQL--TSIROVP--TAAYHRAGIFQAIQNAAEA- 933
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 802 KTLDDYSAPLKEICAIIPDVIMNVLGKAVLVKTLIRSIPTERRKRKSGENSEFKLIAAKAV 861
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 934 -----EQLLPKPGTHSEKSSSESSEANSKDRGLQSNPKTNKRHSAAILPRT 978
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 862 LDKIALDLRPSLNSQFKAAPLAFQSFHTSLVKLFSTQOKIRASQIKGGGGNAGSSSK 921
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 979 GSKGSFVYVIGLTYTSVALLSLITAI-----KSKK 1007
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 922 ECKPFFIHIVSMVITLILAAVLFLAYLCYRKKK 955
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 75
S41552
Probable transcription factor SPT7 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YBR0739; protein YBR081c
C:Species: Saccharomyces cerevisiae
C:Date: 28-Jan-1994 #sequence_revision 09-Sep-1994 #text_change 16-Aug-2004
A:Accession: S41552; S45946; S45948; S40800; S45478; S54985; S59716
R:Gansharoff, L.; Dollard, C.; Tan, P.; Winston, F.
Submitted to the EMBL Data Library, July 1993
A:Reference number: S41552
A:Accession: S41552
A:Molecule type: DNA
A:Residues: 1-1332 <GAN>
A:Cross-references: UNIPROT:P35177; EMBL:L22537; NID:G349189; PIDN:AAC37424.1; PID:G3491
R:Steenma, H.Y.; van der Aart, Q.J.M.
Submitted to the Protein Sequence Database, August 1994
A:Reference number: S45932
A:Accession: S45946
A:Molecule type: DNA
A:Residues: 1-1332 <STE>
A:Cross-references: EMBL:Z35950; NID:G536341; PIDN:CAA85026.1; PID:G536342; MIPS:YBR081c
R:Andre, B.; Criepluch, C.; Hein, C.; Jauniaux, J.C.; Urrestarazu, A.; Visiers, S.
Submitted to the Protein Sequence Database, August 1994
A:Reference number: S45893
A:Accession: S45948
A:Molecule type: DNA
A:Residues: 1-835 <AND>
A:Cross-references: EMBL:Z35950; MIPS:YBR081c
R:Haynes, S.R.; Dollard, C.; Winston, F.; Beck, S.; Trowsdale, J.; Dawid, I.B.
Nucleic Acids Res. 20, 2603, 1992
A:Title: The bromodomain: a conserved sequence found in human, Drosophila and yeast prot
A:Reference number: S40800; MUID:92285152; PMID:1350857
A:Accession: S40800
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 463-523 <HAY>
A:Cross-references: EMBL:M87651; NID:G172683; PIDN:AAA35087.1; PID:G172684
R:van der Aart, Q.J.M.; Barthe, C.; Doignon, F.; Aigle, M.; Crouzet, M.; Steensma, H.Y.
Yeast 10, 959-964, 1994
A:Title: Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyces
A:Reference number: S45462; MUID:95076715; PMID:7985423
A:Accession: S45478
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-624,'LRGKKKI', 633-1332 <VAN>
A:Cross-references: EMBL:X76294
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
R:Gansharoff, L.J.; Dollard, C.; Tan, P.; Winston, F.
Genetics 139, 523-536, 1995
A:Title: The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein important
A:Reference number: S54985; MUID:95229044; PMID:7713415
```


A;Accession: S54985
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1332 <GAW>
A;Cross-references: EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:g349190
R;van der Aart, Q.J.M. Data Library, August 1995
submitted to the EMBL Data Library, August 1995
A;Reference number: S59702
A;Accession: S59716
A;Molecule type: DNA
A;Residues: 1-1332 <VAV>
A;Cross-references: EMBL:X76294; NID:g974203; PIDN:CAA53940.1; PID:e264674; PID:g558360
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SPT7
A;Cross-references: SGD:S0000285; MIPS:YBR081C
A;Map position: 2R
C;Superfamily: bromodomain homology
C;Keywords: nucleus; transcription regulation
F;466-521/Domain: bromodomain homology <BRO>

Query Match 3.4%; Score 173; DB 2; Length 1332;
Best Local Similarity 19.7%; Pred. No. 3.1; Indels 248; Gaps 38;
Matches 165; Conservative 130; Mismatches 296; Indels 248; Gaps 38;

QY 19 THNQEVSLVKEPILKQTOASSISGADYAESGSKLKN-----ETSGPVDDTV 69
DB 615 TNDKLTSLVPEGEKEDKTASST--VTVHENVNKNKEIKNGKNERQDMVESSEKTES 672

QY 70 TDLFSKRTTPEKIKONLAK-----GPREQLKAVTENTSEKQITSGSOLEQSKE----- 120
DB 673 KQDAAKAKOTEDGLQKTAENKEAGENBEEDEDDDEMDVDSQSYLLEKDDDRDDL 732

QY 121 SLSLNKTVPSTSNWEIC---DPIITKG--NTLVGLSKSGVEKLSQTDHLVLPQAAQDGTQ 174
DB 733 EISVWKTIVTAKVRAEICLKRTEYFKNGKLNDSSEAFKPNQRMKRPDQLFLEVY----- 786

QY 175 LIQVSAFPTPKKTAIAEYTSRAGENG-----ETSQLDVGDKETIINEGEVFP 221
DB 787 -----EQKALESVROKIEQNSIMKNGFGTVLQKEDDDQLQFHNHSLNGNEAF 834

QY 222 -----NSVLLKKVTIPGKYHIGQDAFVDNKNI---AEVNLPSLETISDYAFAH 268
DB 835 EKQPNDEILDTRFLQYDISNAIPDIVVEG-VNTKTLQKMEDASVDRMLQNGINKQSRF 893

QY 269 LALKQIDLP-----DNLKAIGELAFDQITGKLSLPLQLM-----FLAERAPKSNHI 316
DB 894 LANKQIDLPKMNQNTLTIQOI-----RHICHKISLIRMLQSPLSAQNRSNPNFALNNHI 949

QY 317 KTIETPRGNSLKVIGEASFQNDLSQLMLPDGLEKIESEAFNGPDGDDHNNRVVLTGSG 376
DB 950 YNYTIIDSLDI-----DPVSQI-----PTHYKNNRELWKFMH 984

QY 377 KNPGLATENTYVNPDKSLWQSPSEI--DYTKWLEEDFTYQKNSVTGFSNKGLOKVRNK 434
DB 985 KNISKVAMANGFETAHPASAINMLTEIAGDYLSNLK--TLKLHHETNSLNRG-----T 1035

QY 435 NLHFKPHNGVITTEIG-----DNAPRV--DFQNKTLKYVDLEEVKLPSTIRKIGAFAPQ 488
DB 1036 NVBMLQ-----TTLENGINRPDLFSYVESEFGKTKKLDQIKQ-KLESFLRALLRPTLQ 1090

QY 489 SNMLKSFASDDLEEKEGAFMNRITETLEKDLVTIGDAAPHINHIYAIVLPSVQBI 548
DB 1091 ELBERNFE--DESQSFTGDPASE-----LTGEDFP----- 1119

QY 549 GRGAFRONGANN-----LIFMGSKVKTL-GEMAFLSNRL--EHLDDLSEQKOLTEI 595
DB 1120 --GFRELGLEKEFGVLSVPLQLLTQFTQFDGETKVQAKKIQPEESDSIVYKKTIG 1176

QY 596 PQVAFSDNALKEVLLPASLKTIREAPKKNHLKQLEVASALSIAFNALDDNDGBOFN 655
DB 1177 MLDA--GSFWNTLLP-----LLQKDYERSKAYIAKQSKSAN-----DK 1213

QY 656 KVVVKTTHNSYALADGEHFI-----VDP-DKLSSTI-----VDLEKIL 692
DB 1214 TSMTSTEDNSFALLEEDQFVSKTKATKARLPPTGKISTTYKKKPIASAFILPEEDLENDV 1273

QY 693 KLIIEGLDYSTLRQTOTQFRDMWTAGKALLSKNSLRQGEKQKFLQBAQFPLGRVLDKA 751
DB 1274 K-----ADPTTTVNAKVGAE--DGDSSSLFLRTPQ-PLDFLDMDDA 1311

RESULT 76
D97033
uncharacterized protein, probably surface-located [imported] - Clostridium acetobutylic
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: D97033
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: D97033
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-1819 <KUR>
A;Cross-references: UNIPROT:Q97K40; GB:AE001437; PIDN:AAK79055.1; PID:g15023996; GSPDB:
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1081

Query Match 3.4%; Score 173; DB 2; Length 1819;
Best Local Similarity 18.3%; Pred. No. 4.9;
Matches 206; Conservative 186; Mismatches 388; Indels 348; Gaps 52;

QY 38 ASSISGADYAESGSKLKNETSGPVDDTVTDLPDSKRTTPEKIKONLAKGPREQ-EL 96
DB 36 SSKKIVTSEKASESGNNAVNNEIK-----TPDNSYEQAKNLQVQADV 79

QY 97 KAVTENTSEKQITSGSOLEQSKESSLNKTVPSTSNVEICDFITKNTLVGLSGSVBK 156
DB 80 KQVQGVTVNNTQ-TSNETDEKLASINAGSENLSYNSLGF-----TNVQPDNIKV 131

QY 157 LSQDHLVLPSQAADGTQLIQVSAFPTPD-----KKTAAEYTSRAGENGSEIS 205
DB 132 MEQIKNARTSGADLTQDEIGTITANTSDSIKIGIFDRMTKGTAVVSDYQFLL-----IT 186

QY 206 QLDVVGKEIINEGEVFNVSLLKKVTIPGKYHIGQDAFVDNKNIAEVNLPSLETISDYA 265
DB 187 QVNSNMLQDINSWLTGKRYALISKILST-----VDKITTAVSNINNANDHESDFT 236

QY 266 FAHLALKQIDLPDNLKAIGELAFDQITGKLSLPLQLMLAERAPKSNHIKTIETFRGNS 325
DB 237 ALQIYVQEDL-----VFFLNENIS-----AAKN-----KQSD 265

QY 326 LKVIAGEASFQNDLSQLMLPDGLEKI-----ESEAFNGPDGDDHNNRVVLTGSKGN 378
DB 266 LSSSEINTVWQDSLKLQ--DALERINLQOGTLDYHFIIGAANVDENLLEDVNGWANGK- 322

QY 379 PSGLATENTYVNPDKSLWQSPSEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVK----- 431
DB 323 --GWFERNAEIBINSIVEPLSRINSRSGDSTAAD--YDAIQVDGVDGDDNNIDNIKSVIAEKK 378

QY 432 --RNKNLEIPK-----OHNGVTITEIGDAPFNVDPONK 463
DB 379 QLKGNLMAIEKSAVEYKTVLDYFDYDLIKGTAKVSDYNAIGLTVTED---NVTDMNE 435

QY 464 TLRYKDLBEV-KLPSTIRKI-----GAFQSNLKSFP 495
DB 436 LLKNRDIKTLNKLQDNINSIIKSLKNINAGTDPDYYTNLNTSSVTQDNISIFRDDIKTA 495

QY 496 EASD--DLEBKEGAFMNRITETLEKDLVTIGDAAPHINHIYAIVLPSVQBIETGRSAF 553
DB 496 KOAGSDLNKGGKIQDSVNSNLSKRLSAMD--INAGSAALDDYNNLGI---EGVT----- 544

hypothetical protein C56E6.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15864
R:Fulton, L.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of *C. elegans* cosmid C56E6.
A:Reference number: S69019
A:Accession: T15864
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1066 <FUT>
A:Cross-references: UNIPROT:Q18902; EMBL:U39996; NID:g1055114; PID:g1055120; PIDN:AAA810
C:Genetics:
A:Gene: C56E6.6
A:Introns: 62/3; 110/3; 376/3; 432/3; 482/3; 574/2; 649/2; 679/2; 769/2; 833/3; 892/3; 9

Query Match 3.4%; Score 172; DB 2; Length 1066;
Best Local Similarity 20.1%; Pred. No. 2.5;
Matches 162; Conservative 129; Mismatches 300; Indels 214; Gaps 39;

QY 5 LKTVLTLTVSVVTHNQEVSLVKEPILKQTOASSISGADYAESSGSKLKINETS 64
DB 212 LKNFAKTLKTL-----KLAENNIHATPEALRDLRLNLNLNGN---KLNKIDGD 258

QY 65 V-----DDTVTDLFSFKRTTPEKIKDLAKGPREQELKAVTENTSEKQITSGSQESKE 120
DB 259 VLKGCDDTLVELFIANNYL-EHPHGVLSQMKLEHLIDISKN-----KIMS---LKKPTS 309

QY 121 SLGLNKTVPST-----SNWEICDFI-----TKGNTLVGLSKSGVEKLSQ 159
DB 310 LLSITKEETSTVRLNLNAGNRINNMDSYLIFFHEHMLLTLYVDSFNRIEISPRVPEKLN 369

QY 160 TDHLVLPQS-----AADGTQLIQVAFPTDKTATAEYTSRAGENG 202
DB 370 LESLFLONQLAHFSLFLDKLRHLMLDNQIKIDNFSLADLPK---LOHLSLAG--- 423

QY 203 ETSQDVGKEIINEGEFVNSYLLKKVTPTGKHK-IGQDAFVDNKNIAEVNLP-ESLET 260
DB 424 --NOLDIITENMPGSS---SSSELKSLNLAHNKIHSISRSFSDLDNLQQLRSLHNNIIT 478

QY 261 ISDYAFAPHLA-LKQIDLPNLKAIGELAPFDNQITGKLSLRLQMLAERAFKSNHIKT 319
DB 479 ITSMTFSNLRNRLRYLDLSH-----RIIKLPSALYQLPALDVL 517

QY 320 EFRGNSLKVIGESFOD-NDLSQLMLPDGLEKIESEAFTCNPGDDHYNRVVLWTKSGKN 378
DB 518 HLDHNNLNEIDRADFRSFDLSQSLKLSHNAFRFSCFELGSIQVHQ----- 564

QY 379 PSGLATENTVWNPDKSLWQESPEID--YTKMLEEDFTYQKNSVTGFSNKGLOKVKRKN 435
DB 565 ---LDLSSNQIN-----EIDIFCTARGIRK-LSLASNSVEKINRKLQDATELTS 610

QY 436 LEIPKOHNGVTITEIGNAF-----RNVDQFQNTLRK-YDLEEVKLPSTIRKIGAFAP- 487
DB 611 IDI--SHNG--IIDVSDAFCECRKLSHKLSHNYIRNLWKTRVCIP-WISHLITFCFF 665

QY 488 -----QSNLKSFE-----ASDDLEIKEGAFNN-NRIETLEKOKL 523
DB 666 TKEHLERTISFSYIIVDSSQLTSFGNLSILSPANNKVDIEDGAFENLLSKILDLSNPN 725

QY 524 VTIGDAAFHNIHIAIVLPESVQIEGRSAFRONGANNLIIFMGSKV--KTLGEMAFLSNRL 581
DB 726 VTWS-----PTAFDLSHSISINMANTGLFSMPKFSHRISQISLNTSCNKI 772

QY 582 EHLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTIREBAFKNHLKOLEV-ASALSHIA 640
DB 773 YELSEKDLAPLTKVVALDISHNNLKQISSNAPEPLI-----HLKQLNVSNPITHLT 824

QY 641 FNALDDNDGDEQFDNKVVKTHNSYALADGEHFIVDDPKLSTIVDLKILKLEGLDY 700
DB 825 -----NEHIQQL-----YKLYNIPD---MARPYQISSILSNLPLHTIYVDIKE 865

QY 701 STL-RQ--TTQTQ-FRDMTTAGKAL 721
DB 866 SALDRQFYTADTRLLRLHLVWAGN 890

RESULT 79
A54817
ATPase ScfII, chromosomal scaffold - chicken
C:Species: *Gallus gallus* (chicken)
C:Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 09-Jul-2004
C:Accession: A54817
R:Saitoh, N.; Goldberg, I.G.; Wood, E.R.; Earnshaw, W.C.
J. Cell Biol. 127, 303-318, 1994
A:Title: ScfII: an abundant chromosome scaffold protein is a member of a family of putative
A:Reference number: A54817; MUID:95014725; PMID:7929577
A:Accession: A54817
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1189 <SAI>
A:Cross-references: UNIPROT:Q90988; GB:X80792; NID:g572691; PIDN:CAA56767.1; PID:g572692
C:Superfamily: chromosome segregation protein SMC1
C:Keywords: nucleotide binding; P-loop
F:32-39/Region: nucleotide-binding motif A (P-loop)

Query Match 3.4%; Score 172; DB 2; Length 1189;
Best Local Similarity 19.7%; Pred. No. 2.9;
Matches 179; Conservative 130; Mismatches 325; Indels 274; Gaps 39;

QY 36 TOASSISGADYAESSGSK-LKINETSGPVDVTDLFSFKRTTPEKIKDLAKGPREQ 94
DB 414 TEAKQAMKLYAQOELKTKQAEVKKMGSYKED-QEAFAIRKTKLEKQDEMCK- 467

QY 95 ELKAVTENTSEKQITSGSQESLSLNKTVSTSNWEICDFITKGNLTVGLSKSGV 154
DB 468 -LK--YEEAEQEAHLAKKKQLSSSEISLR-----ELCEIEAKHPYRFEYKNP 513

QY 155 EK---LSQTDHLVLPQAADGTQLIQVAFPTDKTATAEYTSRAG-NGISQLD 208
DB 514 EKNWNPNCVKGLV-----TLITVKDIS-----TSKALEAVAGGKLYNIV 553

QY 209 VD---GKEIINEGEFVNSYL-LKKVTPTGKHKIGQDAFVDNKNIAEVNLPESLET 261
DB 554 VDTEATGKILEKGQKHKRYTIIPLSKISA---NSIGHEIISLAKNLI----- 598

QY 262 SDYAFAPHLAKQIDLPNLKAIGELAF-----FDNQITGKLSLRLQMLAERAFKS 313
DB 599 -GHRVHTAISLIDYNSLQKAMEYVFTTLVSSMDN--AKVTFDKRIMR----- 647

QY 314 NHIKTIIEFRGNSLKVIGESFODNLSQLMLPDGLEKIESEAFTCNPGDDHY---NNRV 370
DB 648 ---KTVTLQGDIFDPQGTLS---GGASSHVP-----ILSKLTVWRDADELKIKTSQLE 696

QY 371 LWTSGKNPSGLATENTVWNPDKSLWQESPEIDYTKMLEEDFTYQKNSVTGFSNKGLOKV 430
DB 697 ATEKELANLKNMAEYQHLKQWQESSEAEALLQTKIQSSAVHKQOEDL----- 745

QY 431 KKNKLEIPKOHNGVTITEIGNAFRNVDQFQNTLRKVDLEEVKLPSTIRKIG---APAF 487
DB 746 -----LALKK-----TIAECETLTKTESQKAEYEEKALENKNKNAEERKEIKNAQ 795

QY 488 QSNLKSFEASDDLEIKEGAFNNRIETLEKOKLVTIGDAAFHNIHIA---IVLPE 543
DB 796 QKLSNAKKKADDSRKMKE---KQEVVALVLEQLKQEQASYSKQSEAAQAATSLKE 852

QY 544 SVQIEGRSAFR-----QNGANNLI-----FMGSKVKTL-----GEMAFLSNR 580
DB 853 QVSALAEAVKTRSLKNAENELSEKGLMEERTKDIKAKSAKIEKYREQNNELQSLNA 912

QY 581 LEHLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTIREBA---FKNHLKQL----- 630
DB 913 LEH-DINKYQOET-ADASSTLDKLLKEYKWTIASEKELFCQADTTTYDFEANNPKETGQKLQ 970

QY 631 -----EVASALSHIAFNALDDNDGDEQFDNKVVKTHNSYALADGEHFIVDDPKLSS 683

Db 971 KLTTKKLEKLSNRRANLL--SEAEERYNDLMKKR-----WVENDKI-- 1013
Qy 684 TIVDLKILKLEGLD-----YSTLRQTTQTFQDMWTAGKALLSKSNLRQGBKQ 733
Db 1014 -----KILATIEELDRKKNKALHIAWEKVNKDFGIFSMLLPGAKAMLVPS-----KKQ 1062
Qy 734 KFLQBAQFPFLGRVLDLKAIAKAEKALVTMKATKNGQLLERSINKAVLAYNNSAIKKANVK 793
Db 1063 NLLDGLFRVGLGDIWK----- 1079
Qy 794 RLEKELDLTLGLVEGKGPLAQATMVQGVYLLKTPLPPEYIYGLNVFDFKSGKLIYALDM 853
Db 1080 -----ENLTSLGGQSRSLAALSILAILLFK--PAPIYIL-----DEVDAALDL 1121
Qy 854 SDTIGEGQ 861
Db 1122 SHTONIGQ 1129

RESULT 80
D89824
hypothetical protein sdrD [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D89824
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D89824
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1385 <KUR>
A:Cross-references: UNIPROT:Q99W47; GB:BA000018; PID:g13700454; PIDN:BA841751.1; GSPDB:G
A:Experimental source: strain N315
C:Genetics:
A:Gene: sdrD

Query Match 3.4%; Score 172; DB 2; Length 1385;
Best Local Similarity 18.6%; Pred. No. 3.6;
Matches 210; Conservative 150; Mismatches 435; Indels 332; Gaps 47;

Qy 11 TLTIVSVVTHNQEVSLVKEPILKQTOASSISGADYAESSGSKLKI-----NETSGPV 65
Db 29 TVGTASILVGTTLIFGLGNQEAKEASTNKELNEATTSASDNQSSDKVDMQQLNQEDNTK 88

Qy 66 DDTVTDLPSD-----KRTTPEKIKONLAKGPREQELKAVTENTESEK 107
Db 99 NDNQKEMVSSQNETTSNGKSIKESVOSTTGNKVEVSTAKSDQASPKSTNEDLNTKQ 148

Qy 108 QITSGSQLEQS-KESLSLNTKTPVSTSNBEICDFITKGNVLGLS---KSGVEKLSQTD-- 161
Db 149 TISNQEGLQDPLENKSQVNVVQPTNEENKVKDAKTESITLVNKSDAIKSNAETLVNNSN 208

Qy 162 -----HLVPSQAADGTQLIQVAFPTP---DKKTAIAEYTSRAGENCEISQLVDG 211
Db 209 SNNENNADIILPKSTAPKSLNTRMMAAIQPNSTDSKNVNDLITS---NTTLTVVDADN 264

Qy 212 KEIINEGEVFNFSYLLKKVTIPGYKHIGQDAFV-----DNKRIAEVNLPE 256
Db 265 SKTIYPAQ---DYLKLSQITVDDKVKGGDYFTIKYSDTVQVYGLNPEDIKNIGIDKPN 321

Qy 257 SLETTSDVAFALHALKQIDLPNLKAIGELAFDFNQITGKLSLPRQLMLAEAPKSNHI 316
Db 322 NGETI-----ATAKHDTANNLITYFTFDVDRPNSVKMGINSIYNDADTIPVDK-- 371

Qy 317 KTIETFR---GNSLKV-----IGBASFDNDLSQMLPDGLEKIESBAFTGNPGDDHNNRV 369
Db 372 KDVPFVSVTIGNQITTTTADITYPAYKEADNNSI-----GSAFETTVSHVGNVEDPGYYNQV 427

Qy 370 VLWTSKGNPSGLATENTYVNPDKSLMQESPEIDYTKWLBEDFTYQKNSVTGFSNK---G 426
Db 428 V-----YVNP-----MDKDLGAKLKVBAIHPKYPFN 454

Qy 427 LQKYQRN-KNLEIPKQNGVTTITE-----IGDNAFRNVDFQNK 464
Db 455 IGQINQNVNIIKIVRVPBGYTLNKGVDVNTDLVDVTFDFKVKMTYGSNQSVNLDFGDI 514

Qy 465 L-----RKVDLEVKLPSTIRKIGAPAFQSNLKLKSFASDDLEELKEGAPMNRITL 517
Db 515 SAYVMVNTKPYQTNSESPTLQMA---TLSTGKNSVSTGNAL-----GFTNNSQSG-- 563

Qy 518 ELKDKLVTIGDAAPHI-NHIIAIVLPESVQSIGRSAPFRONGANNL---IPMGSKVKTIGE 573
Db 564 -----GAGQEVYKIGNYVWEDTNKNGVQLG---EKGVGNVTVTFDNNNTNPKVE 611

Qy 574 MAPLSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREBAFKKHLKQLEVA 633
Db 612 A-----VTKEGDSYLIPLNPGDYRVFBSNLP---KGYEVTSPSKQGNNEELDSN 657

Qy 634 SALSHIAFNALDDNDGDEQFNDKVVVTHNSYALADGEHFIVDPDKLSSTIVDLKILK 693
Db 658 GLSSVITVNGKDNLSAD-----LGIYKPKYNLGD---YVWEDTNKNGIQDQDE--K 703

Qy 694 LIEGLDYSTLRQTQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIA 753
Db 704 GISGVTV-TLKDNGNVLTQVTTDADG-----KYKF-----TDLNNGNY 741

Qy 754 KAERKAL---VTKATKNGQLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLVEGK 810
Db 742 KVEFTTPEGYTPPTVTSQSDIEKDSNGLT-----TTGVINGAD 779

Qy 811 PLAQATMVQGVYLLKTPLPPEYIYGLNVFDFKSGKLIYALDMSDTTIGEGQKDAYGNPIL 870
Db 780 NM---TLDSGFY--KT---PKYNLGNVW-----EDTNKQKQDSTEKGIS 817

Qy 871 NVDEDSGYHALAVATLADYEGDLIKTLNLSQLTSIRQVPTAAVHRAGIFQAIONAA 930
Db 818 GV-----TVTLKNGEVLQTKTDK-----DGKYQ-----FTGLENGT 851

Qy 931 ABAEQLLPKPGTHSEKSSSSSANSKDRGLQSN-----PKTNRGRHSAILPRTGSK 981
Db 952 YKVEFETPSGYTPQVSGT-----DBGIDSNSTTGVIKDKNDITDSGFYKPTVNL 905

Qy 982 GSFYV-----GILGYT-----SVALLSLITAKKKKY 1008
Db 906 GDYVWEDTNKNGVQDKDEKIGISGVTVTLKDENDKVLKTVTTDENGKY 952

RESULT 81

P82885
hypothetical protein UU482 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: P82885
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: P82885
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4688 <GLA>
A:Cross-references: GB:AB002145; GB:AF222894; NID:G6899476; PIDN:AAF30894.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UU482
A:Genetic code: SGC3

Query Match 3.4%; Score 172; DB 2; Length 4688;
Best Local Similarity 19.4%; Pred. No. 22;
Matches 217; Conservative 173; Mismatches 404; Indels 324; Gaps 55;

QY 14 TVSVTHNQEVF-SLVKEBILK-----QTOASSISGADYABSSGSKLKINETSQP 64
D 1608 SYNDMTNQSISFKIHKPEBEPSPSLTINVDTNNAIKTVAHQAQVHFVHLKTDNDIALE 1667
QY 65 VDDTVTDLF-----SDKETTPEK-----IKDNLAGPREQELKAVTENTSEK-QITSGSQ 114
D 1668 NQOIVEAPFAPTNLNDQKVFAKLNNVNSFNEGELEPNLSGLREETTYRLIKVTFKMK 1727
QY 115 LEQSKESLSLNTVPSTSNW-BICDPITK--GNTLVGLSKSGVEKLSQTDHLVLPSSQAD 171
D 1728 PNKAYELLKNGVIFPEYKNGSQAYEFTTQKFEHKVIDVVSSTNTTQOEITV-----KID 1783
QY 172 GTQ-----LIQVAFAPTPDKKTAIAEYTSRAGENGESQLDVGKE-----IINEGE 219
D 1784 GIQRAWNNKLELVNESILGPOEIKTV-----DNNNSVHLSFDKKEYNLVLNLK 1836
QY 220 VFNYSLLKVTPTGVKHI--GQD-AFVDNKNI--AEVNL-----PESLEITSDYAPAH 268
D 1837 PORRYSLLKINI-----KEVDGQDHEFVKEIINNNSFDVNLQSEITASSVEEINDRA--- 1889
QY 269 LALKQIDLPDNLKAIGELAFFNQITGKLSLPRQLMRLAERAFKSNHIIKTIEFRGNSLKV 328
D 1890 -----PDKL-----NQTIKINLXDD-----NDILKTYNDIATITY-DNEQKV 1925
QY 329 --IGASQDNDLSQLMPLD-----GLEKIE-----SEAF--TGNPGDDHNNRVLMWT 373
D 1926 DAIVKTNAGNQKYLEAIIITNLVFNKNVVIKKIEFNLSQTFIKVGK-----NNTNVIYD 1979
QY 374 KSGKN-----PSGLATENTYVNPDKSLWQSPEDIDYTKWLEEDPT 413
D 1980 ESNLKNLIINDFQIGLSTDVNSTQNIIVANNKHVISLTDLPKVNPHI--SKNLKFKLK 2037
QY 414 YQ-----KNSVTGFSNGK-----LQVKRN-----KN 435
D 2038 FQINGEVVYSPILLTSSIIIVNNKNVINFITDLNLSKNQYRLVDVYIIDNDNDINDKN 2097
QY 436 LEIPKOHNGVTITEIGDNFRNVDFQNTLRKYDLEEVKLPSTIRKIGAFAPQSNLKSFP 495
D 2098 -KVPKANNVTRIIDIAPGK--TTISKSNNTW-----NTTSSQFEFVINSDDGN 2144
QY 496 EASDLE-----EIKEGAFM-----NNR-----IETLEKDKLVTIGDAAFIH 535
D 2145 EYLDNLEATISPKGQGLTLPVKVNIKKONNKYLKQITNLEPENRYV-----2193
QY 536 IYAIVL-----PESVOEIGRS--APRONGANNLIFM--GSKVKTIGEMAFLSNRLEH 583
D 2194 LESILLAPKNTKPLVLEILNKDIFSQTQAGNTKVQIKSQNPSTVDTKQRIKLLKDG 2253
QY 584 LDLSBOKQLTEIPVOAFSDNALKEVLLPASLKTIRBEAFKKNHLEKOLEVASALSHIAFNA 643
D 2254 IQANWNEKQLEITYSA-NDNSTKTAIIKLEKNKLEVEP-----ELTNLEKNRTYFTFKIEL 2308
QY 644 LDNDGDDQDFD-----NKVVVKTTH-----NSYALADGEHFIVDPKLSSTIVLEKIL 692
D 2309 INDDNNTKTPFNKSDSIQDKFIVLSNNQGVGNIIIEIQDRDNNHLSAKIRFELNDLNVL 2368
QY 693 KLIEGLDYSTLRQTOTQPRDMMTAGKALLSKNSLRQGEKQKFOEAOFFFLGRVLDKAI 752
D 2369 -----SNDQQAITYNNQTTSAKVTD-----ONQKYL-EATFSNLVNLKDTII 2412
QY 753 AKAEKALVTKATKNGQLLERSI--NKAVLAYNN-----SAIKKANVXRLE 796
D 2413 NKIEFTKPKNASKINGINDTNVYDATNLIIINDLKITGPLHLTKPEANNKTNIS-VS 2471
QY 797 KELDLLTLGLVEKGPLAQMTVQGVYLLKTPPLPE-----832
D 2472 LEIDTNNHISKNYLPIAKFSDSDGOSVLTNPITAAADKIVTNNNNKELTFNLTLNLSNRQ 2531
QY 833 -----YYIGL-----NVYFDKSGKLIYALDMSDTIGEGQDAYGN---PIINV--- 872
D 2532 YIFKGLYYVNSNQTNIDENKFNKSNVDYKITVPTITTTIQKN--GNWTFPPQNAOKF 2589
QY 873 -----DEDNEGYHALAVATLADYEGLDIKTIILNSKLSQ 905

Db 2590 KFNINSDENVDFSTDLDAITITFSQDHTKTKTITTKLKQ 2627
RESULT 82
S33441
EF protein - Streptococcus suis
C:Species: Streptococcus suis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C:Accession: S33441
R:Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are absent
A:Reference number: S33441
A:Accession: S33441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1822 <SMI>
A:Cross-references: UNIPROT:Q07290; EMBL:X71880; NID:g298031; PIDN:CAA50714.1; PID:g2980
Query Match 3.4%; Score 171.5; DB 2; Length 1822;
Best Local Similarity 19.4%; Pred No. 5.7;
Matches 205; Conservative 143; Mismatches 432; Indels 275; Gaps 41;
QY 16 SVVTHNQEVFLVKEPI---LKQTOASSISGADYABSSGSKLKINETSQPVDDTVTDL 72
D 748 AIEAKRDEAFSKIDDDISLRABQRQAADVAAGAAGDALKELDNKATEAKEKI-----800
QY 73 FSDKRTTPEKIKDNLAGPREQELKAVTENTSEKQITSGSQLESKESLSLNTVPSTS 132
D 801 --DKATTASEINDAKTNG-----EINLDSAEAVGEKAINQAKEKELAKAEVE-NKA-----848
QY 133 NWEICDFITTKGNTLVGLSKS-----GVEKLSQTDHLVLPSSQADGTQLIQVAS 180
D 849 -FEALEKVNPNPILLEEKAYFDIDIKESKEVAVEKINNAEN-----TAEITAAIDRAE 901
QY 181 FAFTPKD-KTAIAEYTSRAGENGESQLDVG-----KEIINEGEVFN 223
D 902 IAYNEDVINAQLDALNKLKESSEETKAAIDANPNLTPPEKAKATAKVEELVNNAE---S 958
QY 224 YLLKVTIPTGYKHIGQDAFVDNKNIAEVNIPESLETISDYAFALHAKQIDLPNL---280
D 959 DILSKPT-PETVQAVEDKA---DKDLAKVEL-----QAAADGAKKGEANPNLTPE 1005
QY 281 -----KAIGELAFPDNQITGKLSLPRQLMRLAERAFKSNHIIKTIEFRGNSLKVIGESAF 334
D 1006 EKVAKKAVEADAVKAVTDAIDKASTPTTEVDTSQGVKA--IDAEFEKATQ-----1054
QY 335 QDNLSQLMLPDGLEKIESEAFNGPDHNNRVVLTGSKGNPSGLATENTYVNPDKS 394
D 1055 -----KDNKNIKAESAESAKAIDDDPNLTPDEKESAKNAVEEAQVATAIDKA 1104
QY 395 LWQSPEDIDYTKWLEEDPTVQKNSVTGFSN-KGLQKV-----RNKNLEIPKOHNGV 445
D 1105 STPDVAQVE-----EDKGVAAILITAKADAKGVIAAKLADEIKKLEKQAEAEKAIDAS 1159
QY 446 TITE-----IGDNAFNVDFQNKTLRKVDLE-----EVKLPSITIRKIGAFQFOSNLK 493
D 1160 TMTNEEKAIAKKALQDVVDKGA-----ELEDAARVATNEIHEATTEK--AKAAELAGEK 1213
QY 494 SF-----EASDDLBEIEGAFMNNRIETLELKDQLVITIGDAAFIHNIHYAIVLPESVOEI 548
D 1214 SLTDTGKEARDAVELAKDKELAKEAIRTEEE-----ATKIVEKLAED 1256
QY 549 GRSAFRONGANNLIFMGSKVKTIGEMAFLSNRLEHLDLSEOKQLTEIPVQAFS---DNAL 605
D 1257 TRKAIEDN-----PNLSDEDKQAEIKKLTDAVAKTLATIRDNAD 1295
QY 606 KEVLLPASLKTIRBEAFKKNHLEKOLEVASALSHIAFNALDD-----NDGD-----EQFD 654
D 1296 KRT-----QEAKEAQAALADLEKAKETQKTADKAAIDRLTLVKGDELEATKQDAK 1345
QY 655 NKVVVKTTHNSYALADGEHFIVDPKLSST-IVDLEKILKLEGLDYSTLRQTQTQ---710

Db 1346 NKIAKAAAEEAIAENPN-LTDAEKTFTDAVDAE-VAKANDAIASAATSPADVQKEEDA 1403
QY 711 -----PRDWTAGKALLSKSNLRQGGKPKQLOAOFFLGRVLD 749
Db 1404 GVAIAEDVDAKQAKNKIAKAAAEEATGSPNLTDAEKTFTDA-----VD 1454
QY 750 KAIKAEKALVTKATKNGCOLLERSINKAVLAYNNSAIKKANVKELEKELDLTGLVGGK 809
Db 1455 AEVAKANDAIASAATSPADVQKEEDAGVAAIAEDVDAKQAKNKIAKESDAKSAIDAN 1514
QY 810 GFLAQTAVQGYLLKTLPLPEYIYGLNVYDKGKLIYALDMSDTIGEGOKDAYGNPI 869
Db 1515 PNLTDAB-----KESAK-KAVDADAKAATDAIDASTSPV 1547
QY 870 LNVEDNNGYHALAVATLADYEGDLKILNSKLSQLSIRQVPTAAVYHRAGIFOAIQNA 929
Db 1548 EAQSAEDKGVGSIQADVL-DAKQDAK--NKIAKEVAAAKEAIDA---NPNLSDAKEA 1600
QY 930 AAEAEQLLPKPGTHSEKSSS--ESANSKDRGLQS 962
Db 1601 SKKAVDADAKATDAIDASTSPVQAQSAEDKGVGS 1635

RESULT 83
D97316
Probable S-layer protein [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: D97316
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A36900; MUID:21359325; PMID:21359325
A:Accession: D97316
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1939 <KUR>
A:Cross-references: UNIPROT:Q97DT3; GB:AE001437; PIDN:AAK81319.1; PID:g15026474; GSPDB:G000000000
A:Experimental source: Clostridium acetobutylicum ATCC9824
C:Genetics:
A:Gene: CAC3389

Query Match 3.4%; Score 171.5; DB 2; Length 1939;
Best Local Similarity 19.1%; Pred. No. 6.3;
Matches 201; Conservative 151; Mismatches 386; Indels 315; Gaps 49;

QY 12 LTTVSVV-----THNQEVSLVKEPILKQTKQASSISGADYAESGSKKLKINETSG-P 64
Db 907 LPTINVTGNSTQWTKNDVVLNVSKT-----AGASGVSKVTVIKDNQTP 949
QY 65 VDDTFTDLFS-----DKRTTPKIKINDLAKQPR-----EQL 96
Db 950 VDIENSSTVYGENGYIFTVINGAGASASTVVVVDK---IDKVKPSVAVSSMKIGEDY 1006
QY 97 -----KAVTESEKQITSGSQLESKESLSLK-----TVPTSNWEICDPIIT 141
Db 1007 NYSATVPAEKQVTNQLQNPNNVAVSGYKYIYSTNGTDMNPIDGTSYTVTSTGN-----AT 1060
QY 142 KGNLTVLGSKSGVEK-LSQTDHLVLPQAADGTOLIQVASFAFTPD----- 186
Db 1061 VNYKFAVNSGVSQVSDVQKQVYIIPKNTV---NNIVTATTSNGADYNGQWTKDNFV 1117
QY 187 -----KKTAAIAYTSRAGENGHISQLDVG---KEIINEGEVFNISYLLKKTPTPTGYKH 238
Db 1118 LSGGISEAFVQYQYSTDPDSDEASWKQGANFELTVSKQDAKYFRAVSKGSGSV 1177
QY 239 QGDAFVDKNNKIAEVNLPSELETISYAFALHAKQIDIDPNLKAIGELAFF----- 289
Db 1178 TSG-----AELKLDKTIPTAMT-----VALKENKTFEPLNTISPNLFFKDTLDVNIY 1224
QY 290 -DNQITG-----KLSLPRQLMRLAERAFKSNHIKTIET 321

Db 1225 ANNDISGIDHYDYOVEDGTGDAYSENGPMKTYSELKLSRQKGV-----YARADK 1277
QY 322 RGNLSKVIAGEASF-QDNDLSQMLPDGLEKTESBAFTGNPGDDHYNRVRVWLVTKSGKRNPS 380
Db 1278 AGNISKVVSSDGFIVDNN-----KATAPKITATAGDKTYNS--WTASDIN-- 1321
QY 381 GLATENT-----YVNPDKSLWQESPEIDYTKWLEBEDFTYQKSVTVQKSNKGLQK- 429
Db 1322 -IVLSNSTAFSDIAGYQYKAGSNGTWTDMP-LSSDGLTKDKKITTINQNTNKLKYFRAVSKS 1379
QY 430 -----VKR-----NKLLEIPKQHGVT-----ITEIGNAFRNVDVFNQKTLRK 467
Db 1380 GYCEENSLMIRRDALAPSTKVDVTGAVNSWTSKSVFVDSNANDNIAPISYIK-IGS 1438
QY 468 YDLBEVKLPSTIRKIGAPAFQSNNLKSFASDD--LBEIKEGAFMNNRIETLELKKDLVT 525
Db 1439 GDMN-KMDSTY--TFNNEINSAVQFKAVSDAGVENV-----SNTIYNVMIDKTTPT 1487
QY 526 IGDAA-----PHINHIYALVLPESVQEIQRSAFRQ-NGANNLI FMGSKVKVTILGEMAFLSNR 580
Db 1488 ITGAAQNGSYSIGRV--IGYNDNFQIETATYKQNGAAQNLTSQSQTTPFGAYS----- 1540
QY 581 LEHLDLSEQKOLTEIPVQAFSDNALKEVLLPASLK---TIREBAFKQHHLKQLEVASAL 636
Db 1541 LEVVDKSGNRSVLNFTLKALP--SVNDVLYTSDSKNAIDAIRNELNTHNDLKE-PYASQI 1597
QY 637 SHIAPNALDDNGDQFDNKVVKTHNSYALADGEHFI VDPDKLSSTVVDLEKILKLIIE 696
Db 1598 -----KASVKALEDRAVALEKE--VMSLKDQTSNIV--KTVDLSLPN 1634
QY 697 GLDYSLTRQTTQTOFRDWTAGKALLSKSNLRQEKQKFLQEAQFFLGRVDLKA----I 752
Db 1635 GKOGIIALENPKMSVYDKIAGG-----SSTLTNSQYAKLINESLYLKDKNLAIKALKDQV 1689
QY 753 AKAEKALVTKATKNGCOLLERSINKAVLAYNNSAIKKANVRLKSKELDLTLGLVEGKPL 812
Db 1690 ARVEAAI-----KALPA--KGSVKKADSVKIKAVNDSYNALNKEQKAL 1730
QY 813 AQATWQGVYLLKTLPLPEYIYGLNVYFKSGKLIYALDMSDTIGEGOKDAYGNP--IL 870
Db 1731 VNPDLVK-----KLNDVVDASKLLHLHNSPNGITVTGVDGTFKSPDVTL 1774
QY 871 NVDEDEGYHALAVATLADYEGDLKILNSKL 903
Db 1775 VVDPKKEGNGQTFNFGADY--VKKATLTNDKL 1805

RESULT 84
T30336
nuclear/mitotic apparatus protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30336
R:Merdes, A.; Ramyar, K.; Vechio, J.D.; Cleveland, D.W.
Cell 87, 447-458, 1996
A:Title: A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assembly
A:Reference number: Z20828; MUID:97053784; PMID:8898198
A:Accession: T30336
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2253 <HER>
A:Cross-references: UNIPROT:P70012; EMBL:Y07624; NID:g1514670; PIDN:CAA68905.1; PID:g1514670
C:Genetics:
A:Gene: NuMA

Query Match 3.4%; Score 171.5; DB 2; Length 2253;
Best Local Similarity 18.8%; Pred. No. 7.9;
Matches 226; Conservative 188; Mismatches 426; Indels 363; Gaps 53;

QY 6 KTVALTUTTSVWTHNQEVFSLVKEPI--LKQTKQASSISGADYAESG-----K 53
Db 125 KQTQELASILRFVLDNEDALCLDDKLIINFLXR-KAPFPSSGNDPSSSSDAMTPNVSHRRK 183

QY 620 EAFKNHLKQLEVASALSHIAPNALDDNDGDEQPDKNVVKVTHNS-----YAL-----AD 670
 Db 619 VCTPKYEIQNEEYPLNGI-PS-----DNEIYTEQKGLDKIEYPLTGVFSD 666
 QY 671 GEHFTVDP-DKLSSTIVDLKILKIEGLDYSTLRQTQTQPRDMTTAGKALLSKSLNRQ 729
 Db 667 NETYITEHIGSLNEICPLNEVL-----LDNEIITEQIGSHDEEYPLNEILPNEVIT- 720
 QY 730 GEKQFLQEAFFLGRVLDKAIAKAEKALVTAKATNGQLLER-----SINKAVLAYN-- 783
 Db 721 -EQDGLDEEYPLTGVFSDN-----ETVITEQKGLDKIEYPLTGVFSDNETYITEHIG 774
 QY 784 -NSAIKANKVRKEKL--DLITGLVKGKPLAQATVQGVYLLK-----TLPPLPEY 833
 Db 775 SLNEICPLNEVLNDNEIITEQIGSLDEAICPLNEILNDNEISTLUKESQNEICTPKQ 834
 QY 834 YIGLVVYFDKSGKL--YALDMSDTIGEGQDAYGNPILNDVDEONEGYHALAVATLAD-- 889
 Db 835 EVSQNEEINTLNEILSDNEIHITEQIGSPNEEY--PLTGVSSDNETYITEHIGSLNEEI 892
 QY 890 -----VEGLDITILMSKLSQLSIRQVPTAAYHRAGIFQAIONAARAEQLLPKPGT 942
 Db 893 CPLNEVLNDNEIITEQIGSNNGMYSLNEV-----PSDNEISTPKQKDSQEBEISTPK--- 945
 QY 943 HSEKSSSESANSKDRGLQSNPKTNRGRHSAILPTGSGKGFVVGILGYS 993
 Db 946 --QKGSQEEINILNEILSDN-----EIHTEQIGSPNEEYPLTGVFS 987

RESULT 86
 JH0720
 tanabin - African clawed frog
 C;Species: Xenopus laevis (African clawed frog)
 C;Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 09-Jul-2004
 C;Accession: JH0720
 R;Hemmati-Brivanlou, A.; Mann, R.W.; Harland, R.M.
 Neuron 9, 417-428, 1992
 A;Title: A protein expressed in the growth cones of embryonic vertebrate neurons defines
 A;Reference number: JH0720; PMID:92398961; PMID:1524825
 A;Accession: JH0720
 A;Molecule type: mRNA
 A;Residues: 1-1744 <HEM>
 A;Cross-references: UNIPROT:Q01550; GB:M99387; NID:g214816; PIDN:AAA49966.1; PID:g214817
 A;Experimental source: tadpole head
 C;Keywords: intermediate filament

Query Match 3.4%; Score 171; DB 2; Length 1744;
 Best Local Similarity 17.7%; Pred. No. 5.7;
 Matches 204; Conservative 203; Mismatches 451; Indels 296; Gaps 50;

QY 17 VVTHNQVPSLVKE--PILKQ--TQASSISGAQVAESGSKLKINETSQVDDTVTDL 72
 Db 656 IISKRKVF-LENEYIPVSKDDLTEFTSHLE--NDSESSQSPDSKLFENKSTEDQLITNL 712
 QY 73 FSDKRTTPEKT-KNLAKPREQELKAVTENTESEKQITSGSQLQSKESLSLNKTVPT 131
 Db 713 ---KSTQENIQSN-----QEHLENLEFDSVDPDT 740
 QY 132 SNWEICDPTTKGNTLVGLSKSGVEKLSQ--TDHLVLPQAAADGTQLIQVASFAPTPDKKT 189
 Db 741 VKFM---YPQENLLLEENVVYGDGLVQMATDENII-NQSSDOLLSDHSHHEETKTS 796
 QY 190 AIAEYTSRAGNGEI---SQLDVDGKEINEGEVPSNLYLLKKVTIPTGYKHIGQDAFVN 246
 Db 797 IAVENHRMSEHAEDVKSSEIPVEISNVSEVBEIIEH--ISDVEEDT-----KQAFDE 848
 QY 247 KNIAEVLNPLESLETISDYAPAHALAKQIDLPNLKAIGELAFFNQITGKLSLPQLMR- 305
 Db 849 RVGEIQINQNESTVLDGVSYSQENSQLEDEVSISEQTEKQPEINEQCLSKSDQIRE 908
 QY 306 -----LAERAFKS--NHKTIETFRGNLSKLVIGEASQDND-----LSQLMLP 345
 Db 909 APTDEVDHVQVDFMQEQSFREVQCLNNIKQEDVDLYNQYDEDSFQNNDEPQELSCDLQ 968

RESULT 87

D96796

probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C;Accession: D96796

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

Query Match	3.4%;	Score 171;	DB 2;	Length 1963;
Best Local Similarity	18.3%;	Pred. No. 6.8;		
Matches 209;	Conservative 156;	Mismatches 391;	Indels 384;	Gaps 51;
Qy	30	EPILKQTQASSSTSGADYAESSGSKLKINETSGPDDTVT-----DLFSDKRITTPK	82	
Db	535	QPAEESTNKSEKVPDTSSENTGE-----VSNPDSSTTVSGESKNPEHNDKNSEK	588	
Qy	83	IKONLAKGPREQELKAVTENTEGEKOITSGSQLESLSLTKT-----VP	129	
Db	589	TVSEVPVNPNEGIVEG-TSQNETEKVPQAEETQTSNGKLANENTGEVSNKPSDSKPVE	647	
Qy	130	STSNWEICDFTK-----GNT-----LVGLSKGVEKLSQTDHLVLPSSQAADG-	172	
Db	648	ESNQPEKNGPATKPNSENGNTTSENGQTEPEKKLELRNVSDIELVSTQNGTVRQHVSLDGI	707	
Qy	173	-----TOLIOVASFAP-----TPDKK-----TAIABYTSRAGENGISEO---	206	
Db	708	PENTDTYFVKVKSASFQDVVYIPVASITEERNGQSVYKITAKAKLQOELNKYVDNFTF	767	
Qy	207	-LDVDGKEIINEGEVNSY--LLKKVT-1PTGYKHG-----QDAFV	244	
Db	768	YLDKKAKE-----ENTNFTSFNLVKANQNPSGTYHLAASLNANEVLGPDPERSYIKDTFT	824	
Qy	245	-----DNKNIAEVLNPESL-ETISDYAPAHALAKQIDLPDLNKAIGELAFDNOQITGK	296	
Db	825	GRLIGKDGKNYAIYNLKELFNLGATVEKLSLKNV-----ALSGK	867	
Qy	297	LSPQLMRILAERAFKSNHIKTTEFRGNSLKVIGEASFQDNDSLQMLPDPGLEKIESEAF	356	
Db	868	-----NDIGSLANEATNGTKIKQVHVDG-----VLAGER-----GVGELLAKAD	906	
Qy	357	TGNPGDDHNNRVVLWTKSGNPSGLATENTYVNPDKSLMOESPEIDYTKWLSDEFTYQK	416	
Db	907	QSGIASSGSFQKGRIV-----NTYETD-----	927	

QY 417 NSVTGFSNGLQKVKGNKLNLEIPKOHNGVITTEIGDNAFRNVDFQNTLRKYDLEEVKLP 476
Db 928 ----AYNIGGLVGHJLTKGNASIAKSAKATVTIS----SNTNRS-----D 962
QY 477 STIRKIGAFAPQSNNLKSFEASDDLEETKE-----GAPMNRRIE-----TLELK 520
Db 963 QTVGGLAGLVQDDAHIQNSYAEGLNNVKKHFGKAGVAGYLVDRTSBEKGIAGELTNVLS 1022
QY 521 DKLVTIGDA--AFH-----INHIIAIVLPESVQIEGRSAFRONG-----ANNL 561
Db 1023 DVNVNNGNAITGYHGTGMKVANTSSKANRVFNVTL-EKDEVSKESPEERGTMLDASQI 1081
QY 562 IPMSKVTIGEMAFNLSRNLHLDLSEKQIETETPVOAFSDN-ALK-----EVLIPASLKT 616
Db 1082 VSKKABINPL-----TLPTVPLSTSGKSDFSKIAHYQANRALVYKNIIEKLLPFFNKS 1136
QY 617 -----IREEAPKKNHLKOLEVASALSHIAFNALDNDGDDEQFDNKKVVVKTTHNSYA- 667
Db 1137 TIVKGNLVKENS-----LYQKELLSAVMMKDDQVITDIVSNKQTANKLLI--HYNDHSS 1190
QY 668 -----LADGEHFIVDPDKLSSTIVDLEKILKLEIGLDYST- 702
Db 1191 EKFLKQYQTFANLAENLNGTGLTYTPNQFLYDRDSI-----VKEVLPLOKLDYQSD 1244
QY 703 -LRQT-----TOTQFRDWTAGKALLSKSNLRQBEKQKFOBAQFFLGRVLDLKAI 752
Db 1245 AIRKTLGISPEVKLTLEYLEQ-----FSKTKQNLGDSLKLLSADAGLAS---DNSV 1294
QY 753 AK--ABKALVTKKATNG-QLLERSIN-----KAVLAYNNSAIKKANVKLEKELD 800
Db 1295 TRGYLVDKIKKKEALLGLTYLEKWNFTNGQVNVKOLVMHPDPFGKNGTSPDLTUIE 1354
QY 801 LLTGLVEGKPLAQATMVQGYLLKTPLPPEYIYGLNVYFDKSGKLIYALDMSDTTGE- 859
Db 1355 L-----GKS-----GFNNLLAKNNVDYTGISLASQHGAT 1383
QY 860 ---GQKDAYGNPILNVDEDNREGYHALAVATLADYEGL--DIKTLNLSKLSQLTISRQVPT 914
Db 1384 DLFSTELHYRKVFLPNTSNMFPKSETKAYIVEEKSTTIEVKT-----KQGLA 1431
QY 915 AAYHRAGIFQAIQNAABAEOL-----LPKPGTHSEKSSSESANSKDRGLQSNPKTNR 968
Db 1432 GTKYISGYVDRTSATWYKRNWVPLPLLTLPERSVFVISTMSLGFAGYDRYRSSDHKAGK 1491

RESULT 89
A26655
myosin heavy chain [similarity] - elime mold (Dictyostelium discoideum)
N/Contains: myosin ATPase (EC 3.6.4.1)
C/Species: Dictyostelium discoideum
C/Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 09-Jul-2004
C/Accession: A26655; A24728; S00250
R/Warick, H.M.; De Lozanne, A.; Levinwand, L.A.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A/Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco
A/Reference number: A26655; MUID:87092266; PMID:3540939
A/Accession: A26655
A/Molecule type: DNA
A/Residues: 1-2116 <WAR>
A/Cross-references: UNIPROT:P08799; GB:M14628; GB:M11938; NID:G167834; PIDN:AAA33227.1;
R/DeLozanne, A.; Lewis, M.; Spudich, J.A.; Levinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A/Reference number: A24728; MUID:86016788; PMID:3901008
A/Accession: A24728
A/Molecule type: mRNA
A/Residues: 2035-2116
R/Wegle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A/Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
A/Reference number: S00250; MUID:88112226; PMID:2828113
A/Accession: S00250
A/Status: nucleic acid sequence not shown

A:Molecule type: DNA

A;Residues: 1734-1893 <WAG>

C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos;

F:1-818/Domain: globular head <HD>

F:89-747/Domain: myosin motor domain homology <MMOT>

F:179-186/Region: nucleotide-binding motif A (P-loop)

F:819-2116/Domain: alpha-helical rod <ROD>

Query Match 3.4%; Score 171; DB 2; Length 2116;

Best Local Similarity 20.08; Pred. No. 7.6;

Matches 213; Conservative 163; Mismatches 373; Indels 316; Gaps 51;

QY 55 KLINETSGPVDD---TVTLFSDKRTTPKIKDNLAGPREQEL-----KAVTENT 103

Db 897 EIRVEDMESELDEKKALENLQNKRSVVEKVRDLBEELQEQKLRNLTLEKIKKKEEEL 956

QY 104 ESEKQITSG-----SOLEQSKESLSLNKTPSPSNWEICDFTKGNLTVLGLSKSGVEKLS 158

Db 957 EEMKRVNDGSDTISRLEKIND--ELQKVEELTE-----SFSESKDKGVLEKTRVRLQS 1010

QY 159 QTDHLV--LPSQAADGTOLI-----QVASFAFTPDKKTATIAEYTSRACENGESIS 205

Db 1011 ELDDLTVRLDSETKDKSELLRQKKLEELKQVQE-ALAAETAAKLAQEAANKKLQSEYT 1069

QY 206 QLDVDGKEIINEGEVFNYSYLLKKVTIPTGYKHI-GQDAFVNDKNIAENVLPESLETISDY 264

Db 1070 ELN-----EKFNSVETARSNVESKKTLESQLVAVNNELDESKONRDALEKKKKA 1119

QY 265 APAHLALKQIDLPNLKAIG-----ELAFDNOITG-----KLS 298

Db 1120 LDAMLE-----EMKQLESSTGEGKSLYDLVKVQESDMEALRNQISELQSTIAKLEKIKST 1175

QY 299 LPROLML-----AERAFKSNHIK-----TIEPRGSLKVIGE-ASFQDNDLSQMLPDG 347

Db 1176 LEGEVALQGELEAEQAKSNVEKQKKVLDLEDKSAQLAEETAAKQALDKLKKLEQE 1235

QY 348 LEKIE--SRAFTGNPGDDHYN-----NRVLMVTKS-----GNPSPGLATENTYV 389

Db 1236 LSEVQTLSEANNKNVNSDSTNKHLETSPNNLLEAEQAKQALEKRLGLSELKHV 1295

QY 390 N---PDKSLWQESPE---IDYTKWLE-----EDPTYQKNSVTGFSNK---GLQKVK--- 431

Db 1296 NEQLEEEKKQKESKRRKVDLEKEVSELKQDIIEEVASKKAVTEAKKQKSELDEIKRQY 1355

QY 432 -----RNKNLETPK--QHNGVTITETGDNAFRNVDFQNTLRK--YDLSE----- 472

Db 1356 ADVVSSRDKSVEQLKTLQAKNEELRNLTAEABQGLDRAERSKKAEFDLEAVKNLEET 1415

QY 473 ---VKLPSTIRK-----IGAPAFQSNLKS--FEASDD--- 500

Db 1416 AKVKAEMKKAETDYRSTKSELDDAKNVSSSEYQVQIKRLNEBELSELSVLEADERCN 1475

QY 501 -----LEEIEGAFPMN-----RIETLE--LKQKLVITG----- 527

Db 1476 SATKAKTASALEKDEIDAANNAKAKAERKSKELEVRVAELESLEDKSGTWNVEFI 1535

QY 528 ---DAAPHINHIYAIVLPESVQIEGRSAFRONGANNLIFMGSKVKTGLGEMAFNLSRLEH- 583

Db 1536 RKQDA--EIDDLRALDRETESRIKSDKDKKNTKQPADLEAKVEEAQREVVVTDRLKK 1593

QY 584 ----LDLSEQ---KQLTEIPVQAFSDNALKEVLL-----PASLKTITREAFKK--NHL 627

Db 1594 LESDIIIDLSTQDQTTETKRIKIEK-SKKLEQTLAERRAAEEGSSKAADEIRKQVQEV 1652

QY 628 KQLEVASALSHIAFNALDD-----NDGDEQFDNKKVVVKTTHNSVALADGEHFIVDP 678

Db 1653 DELRAQDLDSRAALNASEKKISLVAEVDYKQLEDEILAK----- 1694

QY 679 DKL-----SSTIVDLEKILKIE-----GLDYSTLRQTTQTKQ----- 710

Db 1695 DKLVKAKRALVELEEVVDQLEEEEDSRSELEDSKRLTTEVEDIKKKYDAEVEQNTKLD 1754

QY 711 -----FRDMTTAGKALL-SKSNLRQGE--KOKFLQEAQFFLGRVLDL-KAIAKAEKALV 760
Db 1755 EAKKKLTDDVTLTKKQLEDEKKKLNESERAKKRLESENEDFLAKLDAEVKNSRAE--- 1810
QY 761 TKKATNGQLLERSINKAVLAYNNSAIKANVYK-----RLEKELDLTLTGLVEKGFLAQAT 816
Db 1811 -----KORKKYBKLDKDTKYKLNDEAATKTQTEIGAALKEDQIDELRSKLEGE--QAKAT 1863
QY 817 MVQGVVLLTPLPLPEYIYGLNVPYFKSGKLIVALDMSDTIGEGQ 861
Db 1864 QAD-----KSKTLEGEINLRAQIEDEGKIKWLEKEKRALEGE 1903
RESULT 90
G88545
protein F59B2.12 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustli.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88545
A>Status: preliminary
A:Molecule type: DNA
A:Cross-references: UNIPROT:F34487; GB:chr_III; PIDN:CAA77581.1; PID:G3877841; GSPDB:GNO
A:Gene: F59B2.12
A:Map position: 3

Query Match 3.4%; Score 170.5; DB 2; Length 918;
Best Local Similarity 18.6%; Pred. No. 2.3;
Matches 176; Conservative 142; Mismatches 359; Indels 267; Gaps 41;
QY 25 FSLVKEPILKQTAQSSISGADYAESGSKLNETSGPVDVTDVTLFSDKRTTPEKIK 84
Db 55 FAMPKLDASKAAMVHSSSHKHGHSSGSS-----SNTHSLTVVGADGKNITENSEK 107
QY 85 DNLAKGPREQELKAVTENTESEKQITSGSQLESLSLNTKTPSTNWEICDFITKGN 144
Db 108 DGYNK---ESKYDEANENTKIKS--ADGSVIETGK---SHNKSDDASY----- 149
QY 145 TLVLGSKSGVEKLSQTDHVLVPSQAADGTQLIQVAFAPTPDKKTAIAEYTSRAGENG 204
Db 150 -----GLEKSSKT-----YADKNGTMLSNTNKINNQSR 179
QY 205 SOLDVDGKEINEGEVFNLSYLLKVTIPIGYKH-----ICQDAFVDNK 247
Db 180 AALD-EGNEFVNQADGTFLRN-----TGHKNTDEHLSHNVLDNAQMSIGADGTSNI 234
QY 248 NIAEYNLPESLETISDYAFALHAKQIDLPDLNK-----AIGELAFPDNOI----- 293
Db 235 TWRKSGVGDHNAAD--AHNPFESLDQGNKSKQNTSKYKKAASAGSNADPESNLSLK 291
QY 294 -----TGKLSLPLQMLAERAFKSNHIKT-----IEPRGNSLKVIGBASFQD 336
Db 292 NADGTSMSNSTGNFNNTSYDKATAEVMGSKVNVNADGTSSMEASHAGSNSSKINSASQS 351
QY 337 NDLSQLMLPDGLEKLETEAFTGNPGDDHNNRV-VLWTKSGKNVPSGLATENTYVNPDKSL 395
Db 352 SDSL-MVGNIGKISHSTNKTNDYALDEANQASGISSEIQIGKNGORSNLESSIESGRKAE 410
QY 396 WQESPEIDYTKWLEBFTYQKNSVTGFSNKGLOKVRKNKLEIPKOHN-GVTITIEIGNA 454
Db 411 SRNNTAADTLDSDANGTVSSSHSKSASGTSLDE--NHN-----KTHALQASVDEHGNWK 463
QY 455 FRNVD--FQNKTLRKYDLEEVKLPSTIRKIGAFQSN-----NLKSPFASDDL-----E 502

Db 464 NHSIDSGYRNKKTGFEFGNSE--MSASIKNADGTMSQVNVKNDTRNTTYEAEKSALEKKNHE 521
QY 503 EIKEGAF-----MNNRI-----ETLELKDKLVTIGDAAPH 532
Db 522 KNSDGTFKDESKGNSRVRNRDGGSNLAVSGYSGKGGVSSNETTASSNAPNTSADSNQ 581
QY 533 INHIY-----AIVLPESVQIEGRSAFRONGANNLI-----FMGSKV-KTLGEMA----- 575
Db 582 FDHLHQKTANGTEITHAKDSQVAASAKSSLDTSMSAIVDAKGNKVDKTSQAADSHDA 641
QY 576 -----FLSN-----RLEHLDLSEQK-----LTEIPVQAF 600
Db 642 ISASDVEDAKIVKHADRSIESISNDSSNQTAASEHNDSSKQSEHEKRONADGSPFVSSNSA 701
QY 601 SDNALKEVLLPASLTKTIREAFKK-----NHLKOLEVASALSHIAFNALDNDNDGEPDN 655
Db 702 KNAVKDA-----TDVQNAPANVDAAGNSVSEKNSLVENH-----SNDSDASSD 748
QY 656 KVVVTHNSYALADGEHFIVDPDKLS-STIVDLKILKILIEG---LDYSTLRQTQTQTF 711
Db 749 KIHKSVDGT-ELTDAKHSNVSTSHLTGQLAQIQKLMADGSGVMSNDNSHITHEKSR 807
QY 712 RDMITAGKALLSK-----SNLROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKATK 766
Db 808 ADVDAKHKAHLKADGSGSTIDEG-----FKHSDLESRGEQAQK---QRYQK 852
QY 767 NGQLLERSINKAVLAYNNSAIKANVVKLEKELDLTLTGLVEKG 810
Db 853 LONGTESND---VGYEKMAKGGDQTSNHHK---TLAKDGGK 889
RESULT 91
S61535
nucleotide-binding head-stalk protein 183K - Giardia lamblia
C:Species: Giardia lamblia
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S61535
R:Marshall, J.; Holberton, D.V.
J. Cell Sci. 108, 2683-2692, 1995
A:Title: Giardia gene predicts a 183 kDa nucleotide-binding head-stalk protein.
A:Reference number: S61535; MUID:9606038; PMID:7593309
A:Accession: S61535
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1620 <MAR>
A:Cross-references: UNIPROT:Q24984; EMBL:X79815; NID:9871047; PID:9871048

Query Match 3.4%; Score 170.5; DB 2; Length 1620;
Best Local Similarity 19.6%; Pred. No. 5.4;
Matches 228; Conservative 184; Mismatches 426; Indels 325; Gaps 52;
QY 21 NOEVFSL-----VKEPILKQTAQSSISGADYAESGSKLNETSGPVDVT 68
Db 254 NOEVNLTQVSORDEILQDKSMLISTLQNIQVMAEKVDAAQAPFFVARNDLS-----L 308
QY 69 VTDLFSDKRTTPEKIKONLAK-----GPREO-----ELKAVTENTESEKQ- 108
Db 309 AEETIKESVTIEKLMKLSKTSQVAIEQRQDAVELARSVADMOKLSDTVDEKQR 368
QY 109 -----ITSGSQLESKESLSLNTKTPST--SNWEICDFITKGNITLVLGSKSGVEKLSQ 160
Db 369 IVHQLHIDMSWKSHBESISQHKSEIARSAFDESTFIKNAR-----ISKLTTE-LIET 423
QY 161 DHLVLPQAADGTQLIQVAFAPTPDKKTAIAEYTSRAGENGESIQLDVDGKEINEGEV 220
Db 424 Q-----TQI-----ASALEKNOELQSESCKAQAHAEEV-----LLAEREI 457
QY 221 FNSVLLKKVTIPTGYKHIGQDAFVDNKNIAEYNLPESLETISDYAFALHAKQID--LPD 278
Db 458 ANTITQELQATK-----GELEAVRKSNEHQVNSYESLQKTQOSYEALLREKETIUKL 512
QY 279 NLKATGELAFPDNQITGKLSLPLQMLAERA-----FKSNHIKTIFRGNLSKVI 329

Db 513 NLECDNAKAFDDHSEQSTAHYQALREEVERVRHTSENLLQEKMGVINRLTAETIOAIKLD 572
Qy 330 GEASFQDNDLSQMLPDGLEKTESAFTGNCPCDDHNNRVVLTWTKSGKPSG----- 381
Db 573 GERALDSKDTQIAELQKNVSSLMEDL-----TKTSSNSAELSRNL 615
Qy 382 -LATENTYV---NPDKSL-----WQSPEDITYKWLBEEDFTYQKNSVTGFSNKG-LQK 429
Db 616 LEATRKDVQRLESQDEQLRTATNEYKENWEKLDVAPAKDAELSKQTAALHALRDELSQ 675
Qy 430 VGRNKNLEIPK-----QHNGVTITEIGDAPFNVDYQNKTLRKYDLEEVKLPSTIR 480
Db 676 VKNHLLVEVRLLEKVNQAQNTSMTL-----LSNQDEISRLGRELDV--VKNSASIN 726
Qy 481 KIGA-FAPQSNLIKSPFASDDLEETKE--GAPMNRRIETLE-----LKDKLVITGDAAFH 532
Db 727 EYVANKTKQSLNLTISQDELQDSKNEADAALKEKVELDTLRLMDSDSTSTAKAQN 786
Qy 533 INHIVAIVLPESVQBIGRSAPRQNGANNLIFMGSKVKTLGEMAPLSNRLEHLDLSEOKOL 592
Db 787 LQN-----TVDKLQBELSSVS-----SDKLAI SAEMNRVVSELKH-ELSTEXQM 829
Qy 593 TEIP-----VQAFSDNALK-----EVL-LPASLKTIRBE-----AFKQHLKQLEVAS 634
Db 830 REAENSRAQLQISHLEAQVKDAKAKADSEVTRLLSLDTVKEELAIVVDQDAKIAEISR 889
Qy 635 ALSHTAFNALDDNODGDFDNVVKVTHNSYALADGEHFI--VDPDKLSSTIVLEK--- 690
Db 890 KLED--TFPKLQRS--EOTVEVLQAABEKELSVAKLNTDQTIALNDRVANLAAELNKKQN 946
Qy 691 ---ILKLEGLDYSLR---OTTOTQPRDMTTAGKALLSKNLRQGEKQKFLQEAQPLG 744
Db 947 ETEELLAPFKDQYKQKQLEDSRTEVTETATSSGRBEISRLQLQIDNLGEALLQOIBYA 1006
Qy 745 -----RVLDL--KAIAKAEA-----LVTKKATKNGOL----- 770
Db 1007 QKEDALKDOLNSAKAVLLAESAEDKDAIITLTKDOLNLRABELLSSEAKDVTIARYKQDC 1066
Qy 771 -----LERSINKAVLAYN-----NSAIKKNVVRLEKLELDLLTG-LVEGKGP 811
Db 1067 ENLQTSLTSEKKEEAYNLIKQEPAGYKQVSAKQA-----YEAQIASLTGDLAAAKK 1122
Qy 812 LAQATMVQGVLLKTPLPPEYIGLVNVPFKSGKLIYALD-----MSD 855
Db 1123 SEQLEMEIEREMKHASAKQAQELKVTIQKQRELELOKEYLSSQAEALSTNWRLOD 1182
Qy 856 TIGEQKDAYG--NPILNVDEDNEGYHALAVATLADYE-----GLDIKTIL----- 899
Db 1183 BIDKLIQDKGLINQISSLKQE-----ISVSS-ADREHQAKKADKDKITLQDALRIAE 1235
Qy 900 -----NSKLSQ-----LTSIRQVPTAAYHRAGIFQATON--AAAEA 933
Db 1236 DKRVLEBGNKLEBIKQYCECLISEKNKSITELNKKVDRMYNEVIEAQAKNAEIRAA 1295
Qy 934 EQLLPKPGTHSEKSSSESANSK 956
Db 1296 BELTKSQTHLDIVSEKDEKLAK 1318

RESULT 92
S31132
hypochemical protein F59B2.12 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Feb-1995
C/Accession: S31132
R/Sullivan, J.; Du, Z.; Thomas, K.; Wilson, R.; Hillier, L.; Staden, R.; Halloran, N.; G
awkins, T.; Ainscough, R.; Waterston, R.
submitted to the ENBL Data Library, November 1991
A/Description: The C. elegans sequencing project: A beginning.
A/Accession number: S31132
A/Status: preliminary
A/Molecule type: DNA

A;Residues: 1-943 <SUL>
A;Cross-references: EMBL:Z11505
C;Genetics:
A;Introns: 47/1, 132/3, 203/2, 433/3, 544/3, 586/3, 767/3

Query Match 3.3%; Score 170; DB 2; Length 943;
Best Local Similarity 17.8%; Pred. No. 2.5;
Matches 172; Conservative 149; Mismatches 360; Indels 284; Gaps 39;
Qy 25 PSLVKEPILKQTOASSISGADYAESSGSKLKITNETSGPVDVTVTLDFSKRTTPEKIK 84
Db 55 FAMPKLDASKAAMVHSSSHKHGHQSSGSSNTHSLTVVGADG----- 97
Qy 85 DNLAGPPEQELKAVTENTESEKQITSGSQLEQSKESL---SLNKTYTPSTSNWEICDFI- 140
Db 98 -----KNITENSEKKOYKNEKSDDEANETKIKSADGSVAASVQMFVCFPLC 145
Qy 141 -----TKGNTLVGLSKSGVGEKLSQTDHLVLPQAAADGTQLIQVASFAP 183
Db 146 FLKLYTFHPSQIETGKSHNKSDDASSYGLEKSSKT-----Y 183
Qy 184 TPDKKTAAIAYTSRAGENGESIQLDVDGKEIINEGVEFNSYLLKKVTIPTGYKH----- 237
Db 184 ADKNGTMLSNTNKNQNSRQALD--EGNEFVNQONADGTFLRNN---TGKNTDEHLS 238
Qy 238 -----IGQDAFVDKNIAEVLNLPESLETISDYAFALHALKQIDLPDLK----- 281
Db 239 HNVLDENAOQSIGADGTSHTNITNRKSGVGDSDHNAASD---AHSNPFESLDAQNKKSQNYS 295
Qy 282 -----AIGELAFFDNOI-----TGKLSLPRQLMLAERAFKSNHIKT----- 318
Db 296 KKAASASGSHADFPESNLESLEKNADGTSMSNSTGNFNTSYDKATAEEVMSKKNVADGTS 355
Qy 319 ---TEFRGNSLKVTIGEASFQDNDLSQLMLPDGLEKIESEAFPTGNPGDDHNNRV--VLWTK 374
Db 356 SMEASHAGNSKINSASQGSDDL--MVGPNGIKSHSTSNKTDNYALDEANQSAGSISEQ 414
Qy 375 SGKPSGLATENTYVDPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRNK 434
Db 415 IGRKQBSLNESSIESGRKAFSRNNTAADTLDSVDANGTVSSSHSKSAGSTSLDE---NH 471
Qy 435 NLEIPKQHN--GVITTEIGDAPFNVD--FONKTLRKYDLEEVKLPSTIRKIGAFAPQSN- 490
Db 472 N-----KTHALQASVDEHGNKKNHSIDGSYNKKTGFEQNSE--MSASIKNADGTMSQVNV 525
Qy 491 ---NLKSFEASDDL-----BEIKEGAF-----MNNRI----- 514
Db 526 KNDTNRTYEAEKSALEKHEKNKNSDGTFKDES KGSNSRNVNRTDGSNLA VGSYSVGKGV 585
Qy 515 ---ETLEKDKLVIGDAAFHNIHY-----AIVLPESVQEI GRSAFRQNGANNLI-- 562
Db 586 SSNETIASNAFNSTDAAESNQFDHLQKTANGTEITHAKDSKQVAASNAKSSLDTSMA 645
Qy 563 --FMGSKV-KTLGEMA-----PLSN-----RLEHLDLSEQ 589
Db 646 VDAKGNKVDKTSQASDASHDAISASSVDVAKIVKHADRSSESINDSSNQRTASEHNDSSKQ 705
Qy 590 KQ-----LTEIPVQAFSDNALKEVLLPASLKTIRBEAFKK-----NHLKQLEVAS 634
Db 706 SEHEKRNQADGSPSDVSSNSAKKNAVKDA-----TDVRQNAFANVDAAGNSVEVKNSL 759
Qy 635 ALSHIAFNALDDNDCGDFQFNKVVVTKTHNSYALADGEHFI VDPDKLS-STIVDEKILK 693
Db 760 VENH-----SNDSDASSDSKIHKSVDGT-ELTDAGHSNVSTSHLTGEOQAIOIKKLA 811
Qy 694 LIEG---LDYSTLRQTQTQFRDMTTAGKALLSK-----SNLRQGEKQKFLQEAQFPLGR 745
Db 812 MADGSVMSNDNSHITHEKSRADVDAAGHKAHLSKADSGSTTDIDEG-----FKHH 860
Qy 746 VDLDKAIKAKFALVTTKKATKNGQLLERSINKAVLAYNNSAIKKNVVRLEKLELDLTGL 805
Db 861 SDLESRGEGAKQ---QRYOKLGNGTSESSMD---VGYEKSMAGKGGDQTSHHK---TLA 909

Qy	676	VDPKLGSTIV---DLEKILKLIEGLD---YSTLRQTQTQPRDMTTAGKALLSKSNLRQ	729
Dy	:	:	:
Db	615	EAEAAAKEALKNDVNNAYEIVQGADGRYYVLK-----I EVADEEEEP	658
Qy	730	GEKQKFQIEAQOFFIGRVDLDKAIAKAEXALVTTKATKNGQLLERSINKAVLAYNNSAIKK	789
Db	:	:	:
Db	659	GEDTPVEQS-----GYATYEBAEAAKEA-----LKEDKVNNAYEVVVQGDGRY	702
Qy	790	ANVKRLKEKELDLLTGLVGEGKPLAQTWVQGV-----YLLKTP-----LPLPEYIIGLVNY	840
Db	:	:	:
Db	703	YVLUKI EDEQE QGEEPCENP-----GITDEWLLKNAKEDA I K E L K E A G I S S D I Y	754
Qy	841	FD--KSGKLIYALD-MSPTIGEGOKDAYG-NPILNVDE-----	874
Db	:	:	:
Db	755	FDAINKAKTVGVEALKNELLKAHA EKPGENPGITIDEWLLKNAKAEA I K E L K E A G I T A B	814
Qy	875	-----DNEGYHALAVATL---ADYEG-----LDIKTILMSKLSQLTSIRO--VP	913
Db	:	:	:
Db	815	YLFLNLINKAKTVGESLKNILKAHAEKPGENPGITIDEWLLKNAKEDA I K E L K E A G I T	874
Qy	914	TAAVHRA-----GIFQAIONAAAEEQLLPKG-----THSEKSSSESANSKD	957
Db	:	:	:
Db	875	SDIYFDAINKAKTTIEGV-EALUNELLKAHKDEEPKPKGEDDKPKDGDDKPKEDKK	933
Qy	958	RGLQSNP-----KTNRG-----RHSAILPRTGSGSVFYVILGYTSVALLSLITA-----	100
Db	:	:	:
Db	934	PGEDKKPEDKKPGKGTDXDSPNKKKKAKLPKAGSBAEIL-----TLAAAAALSTRAAGAYV	986
Qy	1003	-IKKKK 1007	
Db	:	:	:
Db	987	SLKKRK 992	
RESULT 94			
F86378			
protein F21J9.12 [imported] - Arabidopsis thaliana			
C;Species: Arabidopsis thaliana (mouse-ear cress)			
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004			
C;Accession: F86378			
R;Theologis, A.; Eckert, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.			
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.;			
ansen, N.F.; Hughes, B.; Huizar, L.			
Nature 408, 816-820, 2000			
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.			
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.;			
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.			
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.			
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.			
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.			
A;Reference number: A86141; PMID:21016719; PMID:11130712			
A;Accession: F86378			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1-1864 <STO>			
A;Cross-references: UNIPROT:Q9FYL7; GB:AE005172; NID:g9743334; PIDN:AAF97958.1			
C;Genetics:			
A;Gene: F21J9.12			
A;Map position: 1			
 Query Match 3.3%; Score 170; DB 2; Length 1864;			
Best Local Similarity 19.2%; Pred. No. 7;			
Matches 222; Conservative 176; Mismatches 401; Indels 356; Gaps			
Qy	23	EVPSLVKEPIKQTOASSISTGADYASGGKS-----KUKINETSGPVDDTVTDLES	74
Db	:	:	:
Db	333	EFEKLAELAELEKTKCTNTKEKLSMAVTKGALVQNRDALKHQLSSEKTTTELANRLTEL-Q	391
Qy	75	DKRTPTPKIDNLA KGPREQELKAVTENT-BSEKQITSGSQEQSKESLSLNK-----T	127
Db	:	:	:
Db	392	EKEALE--SSEVMWGQLEQSL-----TEKTBDELEKYAEINDRKSVSELYELTKLEQSL	446
Qy	128	VPSTSNWEICDFITKGNLT-LVGLSKSGVEK--LSQTDHLVLP SQADGTQLTIQVASFAFT	184

Db 447 AEKTKELSEC--LTKQEMSTALDQSELDKELAKSDAMVASYQ-----EMLSVN-SII 498
Qy 185 PDKTAIAE-YTSRAGENGESIQ-----LDVPGKEIINEGVFNFSYLLKKVTPTGYKHIG 239
Db 499 ENIETILSNIVTPBEGHSDIVEKVRSLAERKELTNVSYQVNR--LKDL----- 546
Qy 240 QDAFVNDKNIABVNLPSLEIIS-DYAFALALAKQIDLPDLNKAIGELAFDNIQTG-KL 297
Db 547 -----IVSIDLPEMSOSSLESRLAWLRESFLOGKDEVNL-----QNRIESVSM 591
Qy 298 SLPRQLMELABAFKSNHKTIEPRGNSLKVIGEASPDNDLSQLMDGLKELKESAEFT 357
Db 592 SLSEAE-----EKSNIKELDDLSF-----SLKMEETAEAR 623
Qy 358 QNPDDHYNNRVVLWTKSGKNPSGLATENTYVNDPKSLWQSPSIDYTKWLEEDFTYQKN 417
Db 624 GSLEREIEVRLV-----ETSGLWTEGV-----EDHT--SS 652
Qy 418 SVTGESNGLQKVKRNKLEIPKQHGNGVITIEIGNAPRVND-----FQNTLRKYVDLEEV 473
Db 653 DINLVDRSFQKIEKQ-----IRDSSDSSYGNEEIEFAFQS--LLYVRDLFS 698
Qy 474 KLPSTIRKIGAFAPSNLKGFEASDDLE-BIKEGAFMNNRIETLE-----LK 520
Db 699 LCKEMLGEGELISFQVSNL-----SDELKIASQELAFVKEEKIALKDLERSEKSAALLR 753
Qy 521 DKL---VTIGDAAFHNNIYIAVLPESVQIEGRSAFR-----QNGANNIIFMGSK--- 567
Db 754 DKLSMAIKKGKGLVQREKFKTQLDKKEKSEIKLMLEQLQGGTVGVKYNQIDMLSRDLE 813
Qy 568 -----VKTGEMAFSLNLEHLDSLSEK-----OLTEIPVQAFSDNALKEV----- 608
Db 814 RTKLETELATKBERDQLOQSLSLIDTLQKWKSVSEIIALPVDLASEDPSEKIDRLAG 873
Qy 609 -LLPASLKTIRE-EAFKK-----NHLKQELVASALSHAFNALDDGDSQFQNKV 657
Db 874 YIQEVQIARVQEIEKVEKSEVDALTSLKLAETQTALKLVEDALSTADNLSRLTEENRN 933
Qy 658 VVKTHNS-----YALADGEHFIYDPDKLSSTIVDLKIL----- 692
Db 934 VQAKENAELELOKAVADASSVASSELDEVLATKSTLEAALMQAERNISDIIEKEEAQGR 993
Qy 693 -----KLIEGLDYSTLRQTQT-----QFRD---MTT 716
Db 994 TATAMEQEMLQKEASIQKNKLTEA--HSTINSLEETLAQTESNMDSLSKQIEDDKVLTT 1051
Qy 717 AGKALLSKSNLRQSKOKFLOBAQPFGLGRVLDLDAIAKAEKAL----- 759
Db 1052 SLKNELEKLTAEAFERNKMAEASLTI--VSHEEALMKAEKNSLSALQEMKARGESTL 1109
Qy 760 -----VTKATKNGQLLERSI-----NKAVLAYNNSAIKKAN-----VKRLEKBL 799
Db 1110 SSKLVNCHIELAGSGNSQSLSLEIITHLDNLMQLDKDGLISKVNEFLQKPKSL-RDV 1168
Qy 800 DLL-----TGLVECKGFLAQATWVGQVYLLKTPPLPEYIYGLNVFVPSGKLIY 849
Db 1169 DVIAIDITRNTGENGLLAGEMGNAEAVLL--ITLL-----YFQDDSTEAK 1211
Qy 850 AL--DMSDTTIGEGQKDAYGNPILNVDEN-----EGYHALAVATLADYEGL--DIK 896
Db 1212 SLLSDLSNVTEPENSQGS-----ADEDEISSLRKMAEGVRLNKTLENNPEGFSTSID 1268
Qy 897 TLNLSKLSQLTSIR--QVPTAAYHRAGI---FOATONAAAEABQLPKPGTHSEKSSSES 952
Db 1269 TLIATLMQNTAARADVLNIVGHNSLSLEEQVRSVENIVREQENTI--SALQKDLSSLISA 1326
Qy 953 ANSKDQGLQSNPKTN 967
Db 1327 CGAAARELOLEVKNN 1341

RESULT 95

S46773
myosin heavy chain - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YHR023w; type II myosin
C:Species: Saccharomyces cerevisiae
C>Date: 07-Oct-1994 #sequence revision 11-Aug-1995 #text_change 09-Jul-2004
C:Accession: S46773; S05806; S12323
R:Du, Z.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 8082.
A:Reference number: S46773
A:Accession: S46773
A:Molecule type: DNA
A:Residues: 1-1928 <DUZ>
A:Cross-references: UNIPROT:P08964; EMBL:U10399; NID:G500689; PIDN:AAB68872.1; PID:G50006
R:Watts, F.Z.; Shiels, G.; Orr, E.
EMBO J. 6, 3499-3505, 1987
A:Title: The yeast MYO1 gene encoding a myosin-like protein required for cell division.
A:Reference number: S05806; MUID:88111539; PMID:3322809
A:Accession: S05806
A:Molecule type: DNA
A:Residues: 1-45, 'T', 47-58, 'S', 60-329, 331-342, 'S', 344-528, 'ARGHNR', 536-540, 'V', 542-598, 61
A:Cross-references: EMBL:X06187; NID:G4015; PIDN:CAA29550.1; PID:G4016
R:Sweeney, F.P.; Watts, F.Z.; Pocklington, M.J.; Orr, E.
Nucleic Acids Res. 18, 7147, 1990
A:Title: The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide sequence.
A:Reference number: S12323; MUID:91088308; PMID:2263482
A:Accession: S12323
A:Status: translation not shown; significant sequence differences
A:Molecule type: DNA
A:Cross-references: EMBL:X53947
C:Genetics:
A:Gene: SGD:MYO1
A:Cross-references: SGD:S0001065; MIPS:YHR023w
A:Map position: 8R
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; nucleotide binding; P-loop
F:78-779/Domain: myosin motor domain homology <MMOT>
F:180-187/Region: nucleotide-binding motif A (P-loop)
F:459-528/Domain: actin binding #status predicted <ACB>
F:186/Binding site: ATP (Lys) #status predicted
Query Match 3.3%; Score 170; DB 2; Length 1928;
Best Local Similarity 20.0%; Pred. No. 7.3;
Matches 200; Conservative 150; Mismatches 349; Indels 302; Gaps 47;
Qy 57 KINTSGFVDVTVDLFSDKRTTPEKIKDNLAKPREQELKAVTENTESEKQITSGSOLE 116
Db 865 KFNEQINKNLNDLQEMESKKKFLSEK-----NQTVNELENTQ-----DLIN 906
Qy 117 QSKESLSLN-----KTVFSTSNWEICDFTIKNTLVGLSKSGVEKLSQTDHLVLSQA 169
Db 907 QEKENLRKNSLNRVTSSTETLQKQFDLVSEKDEI-----SREKLEVAQNLEEAHQK 960
Qy 170 ADGTQLIQVASFAPTPDKTAIAEYTSRAGE-NGEISQLDQV-----GKEIINEGVFN 222
Db 961 IQGLQ-----ETIRREATLEKLSKNNELIKQISDLNCDISKSSQSLIKESKLKL 1013
Qy 223 SYLLKKVTIPTGYKHIGQDAFVDNKNIAEYNLPESLET-----ISDYAFALHAKQIDL 276
Db 1014 ENETKRLKDVINSKEEIKSFNDKLSSEEDLDIKLVLEKNCNIAMRSLQSLVTENS DL 1073
Qy 277 ---PDNLKATGELAFPDNQITGKLSLPQLMRLAER----- 309
Db 1074 RSKNENPKK--EKAALNNQLNKES---ELLKMKEKIDNHHKELATTSFKQRDDAVSBHGK 1128
Qy 310 ---AFKSNHKTIEFRGNSLKVIGE-ASFO-----DNDLSQLMDGLKELKISEA 355
Db 1129 ITAELKETRQLTEYKSNYQKIEEYSNFQREYKEQKKNLSLVESINDSKILEARL 1188
Qy 356 FTGNPGDDHYNNRVVLWTKSGKNPSGLATENTYVNDPKSLWQSPSID----- 403
Db 1189 SQEISLNOYLNKRI-----SGNSVETNISSTRSTSYSDPDLDKEDIKKYYDLQ 1238


```
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1881 <KUR>
A:Cross-references: UNIPROT:Q9L7Q2; GB:AE005672; PIDN:AAK74809.1; PID:g14972138; GSPDB:Q
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0664

Query Match      3.3%; Score 169.5; DB 2; Length 1881;
Best Local Similarity 19.2%; Pred. No. 7.5;
Matches 220; Conservative 176; Mismatches 443; Indels 309; Gaps 53;

Qy 29 KEPILK-----QTQASSISGADYAESGKS-----KLKINETSVPV 65
Db 388 KEPVDKSELNQIDKASSVSPDYSTASYNALGPVLETAAGVYASEPVKQPEVNSETNKL 447
Qy 66 DDTVDFLSDKRRTPPEKIKDNLAKPRE-----QELKAVTENTSESEKQITSGSQLEQSK 119
Db 448 KTAIDALNVDTKELTNNTIADAKTKVKEHYSDRSWQNLQ--TEVTKAEK-VAANTDAKQSE 504
Qy 120 ELSLUNKTVPTSNWEICDFITKQNTVLGLSKGVKEKLSQTDHLVLPQQAADGTQLIOVA 179
Db 505 VNEAVEKLTATIEK-----LVLESEKPILTSTDKKILERE-----VA 544
Qy 180 SPAFTPKKTAIAEYTSRAGENGELISQLDVGKEIINEGEVFNSEVLLKKVTIPTGYKHIG 239
Db 545 KYTLENQNKTKIKSITA-----ELKK-----GESEVINT-VLTTDDKVTTEISAAPKNLE 593
Qy 240 --QDAFVQNKIAEYNLPESLETISDYAFAPHLAKQIDLPNLKKAIGELAFDFDQITGKL 297
Db 594 YKKEYTLSTWYDRNGEETLENQNI-QLDLKKVEL-KNIKRT-DLIKYEN---QKE 647
Qy 298 SLPROLMRLAERAFKSNHKTIEFRGNLSKVIGEASFQDNLSQLMPDGLKEIESEAPT 357
Db 648 TNESLITTPDD--KSNYY-----LKIT-----SNNQKTTLL--AVKNIESETTVN 688
Qy 358 GNPQDDHYNRVLTGSKGNPSGLATENTYVNPDKSLWQSPEDIDYTKWLEEDFTYQKN 417
Db 689 GTP-----VYKVTADNLVSTADNKP-----EESVHYIEKPKVHEDN 728
Qy 418 SVTGF-----SNGLQKVKRNKNLEIPKQHNQ-----VTITEI 450
Db 729 VYVNFELVEALQNDPSKEYRLQGSMSARNVVPNGKSYITKEFTKLLSSGKQPAITEL 788
Qy 451 GDNAF-----RNVDFONTKRYDLEEV-KLPSTIR-----KIGAPAFQSNMLK 493
Db 789 EHPFNVTNATINNVPENVEIERSQDNFIASLANTMKGSSVITNVKITGTLGRNVA 848
Qy 494 SPEASDDLEEIKEGAFMNRITELKDKLVITIGDAAPHINHIYAIVLPESVQIEGRSAF 553
Db 849 GF-----VNNMNDG-----TRIENVAFFGLKHSNGSGHTGGIAG-----TYRGIVRKAY 895
Qy 554 RQNGANNLIFMGSKVKT-----LCEMAFLSNRL--EHLDLSEKQQLTEI 595
Db 896 -----VDATIGNKTRASLLVPKVDYGLTDLHLICTKALLTESVVGKIDVSNPVEVCAI 950
Qy 596 -----PVOAFSDNALKEVLLPASLKTIR-EEAFKQNHKLQLEVASALSHIAFNALDDNDG 649
Db 951 ASKTWPGVTGNSV-----SYAKIIRGEELFGSNDVDDSDYASAHIKDLYAVEGYSSG 1003
Qy 650 DEQPDNKVVVTKHNSYALADGEHEFIVDPDKLSSITVLEKILKLI EGLDYSTLRQTTQT 709
Db 1004 NRSFRKSKTFTKLTKEQADAKVTTFNITADKLES---DLSPLAKLNEKAYSSI-QDYNA 1059
Qy 710 QPRDMTTAGKALLSKSN-----LRQCEKQFLQEAQFFLGRVDLDKAIKAEKALVTKKAT 765
Db 1060 EYNQAYKNLEKLI PPYNKYDVIYQGNK-----LNKEHLNTEKVLVSVTAM 1104
Qy 766 KNGQLLER--SINKAVLAY-----NNSAIKKANVXR-----LEKEL 799
Db 1105 NNNEFITNLDEANKIIVHADGTQDYFNLSSESSEGLSNVKEVTITDLGIKVTNPVQKD- 1163
Qy 800 DLLTGLVEGKPLAQATVQGVYLLKTPLPPEYYIGL--NVYPPDKS-----GKLIY 849

Db 1164 --NTTLVNDIKSILESVELSQTYQHNLRLGDTVRVNAIKDLYLEESFTDVKENLTNLIT 1221
Qy 850 ALDMSDTIGEGQKDAYGNPILNVDEDNEGHYHALAVATLADYEG-----LDIKTIL----- 899
Db 1222 KLVQNEEBHQLNDSPAARQMIRDKVEKNKAALLGLTLYLNRYYGKVGFDVNIKELMLFKPD 1281
Qy 900 --NSKLSQLSIIRQVPT-----AAHYRAGIFQAIONAAEA----- 933
Db 1282 FYGKSVSLDRLIBIGSKENNIKIGSRTFDAFGQVLAKYTKSGNLDLAFNLNRLQFTNIDN 1341
Qy 934 --EQLLPKPGTH---SKSSS--SESANSKORGLQSNPKTNRGRHSAILPRTGSGKSFYVG 987
Db 1342 MNDWFIDATEDHVIAERASEVEEIKSKHRAP--DNLKRSRLRNT-ILPLLNIDKAHLYL 1399
Qy 988 ILGYTSA 995
Db 1400 ISNYNAIA 1407

RESULT 98
TO1362
probable myosin heavy chain At2g34730 - Arabidopsis thaliana
N:Alternate names: hypothetical protein T29F13.6
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: TO1362; C84760
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome II BAC T29F13 genomic sequence.
A:Reference number: Z1479
A:Accession: TO1362
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-829 <ROU>
A:Cross-references: UNIPROT:O64584; EMBL:AC003096; NID:g3132469; PID:g3132472
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84760
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-829 <STO>
A:Cross-references: GB:AE002093; NID:g3132472; PIDN:AAC16261.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g34730; T29F13.6
A:Map position: 2
A:Introns: 728/2; 770/3

Query Match      3.3%; Score 169; DB 2; Length 829;
Best Local Similarity 18.4%; Pred. No. 2.3;
Matches 177; Conservative 155; Mismatches 328; Indels 304; Gaps 45;

Qy 28 VKPEILKQTOASSISGADYAESGSKLKN-----ETSGFVDDTVTDL-----FSDKR 77
Db 1 MEERVVKSENGS-----LEFHDDTLSSSLQVNGVLKENENPDVDFLEDLSYMEDINDRL 55
Qy 78 TTPKIKDNLAKPREBELKAVTENTSESEKQITSGSQLEQSKLSL-----NKTVPSTSN 133
Db 56 TISRVSDSIIRG-----MVTAEISDAEAKIAQDXLELSKIRETLLLYHYGVSENESES 110
Qy 134 WEICDFITKQNTVLGLSKGVKEKLSQTDHLVLPQQAADGTQLIOVASFAFTPDKKTAAE 193
Db 111 RLHDEUTQG-----SSSLKKKARQKQLMLVEELTNRLREYTHINGSATVD----- 157
Qy 194 YTSRAGENISQLDVGKEIINEGEVFNSEYL--LKKVTIPTGYKHIGQDAFVQNKIAE 251
Db 158 -----DSLGLDSSPHETRSKTVDKMLDSLKSIL-----ETVLKRN--D 194
```


CjDate: 12-Oct-1989 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
CjAccession: A32491
RjGeorge, E.L.; Ober, M.B.; Emerson Jr., C.P.
Mol. Cell. Biol. 9, 2957-2974, 1989
A>Title: Functional domains of the Drosophila melanogaster muscle myosin heavy-chain gene
AjReference number: A32491; MUID:89384556; PMID:2506434
AjAccession: A32491
A>Status: Preliminary
AjMolecule type: DNA; mRNA
AjResidues: 1-2385 <GEO>
AjCross-references: UNIPROT:P05661; GB:M61229; GB:M27194
AjNote: the authors translated the codon TGC for residue 329 as Ser
CjGenetics:
AjGene: FlyBase:Mhc
AjCross-references: FlyBase:FBgn0002741
CjSuperfamily: myosin heavy chain; myosin motor domain homology
CjKeywords: ATP; nucleotide binding; P-loop
Fj137-1032/Domain: myosin motor domain homology #status atypical <MMO>
Fj227-234/Region: nucleotide-binding motif A (P-loop)

Query Match 3.3%; Score 169; DB 2; Length 2385;
Best Local Similarity 20.0%; Pred. No. 11;
Matches 216; Conservative 154; Mismatches 418; Indels 292; Gaps 47;

QY 10 LTLTVSVVTHNOEVSIVKBPILKQTOASSISG--ADYAESGCKSKLKINETS GPVDD 67
DB 1483 LTQEAVALERNKKE---LEQTIQRKDKELSSITAKLSEQVVVLKHQRIKELQARIEE 1539
QY 68 TVTDLFADK--RTTPEKIDNLAKPRQEKLKAVTENTESKQITSGSLQESK----ES 121
DB 1540 LEEVEAEERQAKAKQADLAR-----ELEELGERLEEAGGATS-AQIELNKKREAE 1593
QY 122 LSLNKTVPSTSNWEICDFTIKGNTLVLSK-----SGVEKLSQTDHLVLPQAAADGTQLIQ 177
DB 1594 SKLRDLLEAN-----IQHESTLANLRKHNDVAEAMAEQVDQ----- 1632
QY 178 VASFAFTPDKKTAIAEYTSRAGENSEISQ-----LQVDCKEIINEGEVENSYLLKKVTIPT 233
DB 1633 -----NKLKAKAEHRDQCHN-ELNQTRTACQLGRDKAEKEK--NEYYGQNDLRA 1681
QY 234 GYKHTGQDAFVDNK-----NIAEVLNPESELTISDYAFALHAKQIDLPNLKAI 284
DB 1682 GVDHITNEKAAQEKIAKQLOHTLNEVQSKLDNETNLDPDASKKL-SIENSLRLQLE 1740
QY 285 ELAFPDNQITG-KLSLPRQL---MRLAERAFKSNHKTIEFRGNSLKVIGASPDNDLS 340
DB 1741 EAEQVSQLSKIKISLTQLEDTKRLADESRER-----ATLLGKFRNLEHLD 1789
QY 341 QLMPLDGLKTESRAFTGNPGDDHYNRVVLWTKSGKNPSGLATENTYVNPDKSLMQ--- 397
DB 1790 NLK-----EQVEEEA-----EGK--ADLQRLSKANAEQVWRKY 1823
QY 398 ESPIDYTKWLEE-----DFTYKNSVTGFSNK--GLQKVKR-----NKNLEIPKQ 441
DB 1824 ESDGVARSEELAEAKKLQARLAEABETIESLNQKICIGLEKTKQRLSTEVEDLQLEVDR 1883
QY 442 HNGVTITTEIGNAFRV-----DFQNTKRLKYDLEEVKLPSTIRKIGAPA 486
DB 1884 NAIANAEEKKKQAFDKIIGEMKLVDDLAELDASQKECRNYSTELFRK-----GAYE 1937
QY 487 PQSNLKSFEASDLEELKEGAFPMNRITETLELKDVLVIGDAAPHINHIYAVILVPESVQ 546
DB 1938 -----EQEQLEAVRR-----ENKNLADEVKDLDDQIGEGG-----RNIIH 1972
QY 547 EIGRSAPRQNGANNLIFMGSKVTIGEMAFLSNRLHLDLSEK-----QLTEIPVQAFSD 602
DB 1973 EIEKARKLEAKDEL-----QAALAEAEAELEQENKVLRAQLSELQVQRQID 2021
QY 603 NALKEVLLPASLKTIRREAF---KQHLKQLEVASALSIAFNALDDNDGDGPDKVWV 659
DB 2022 RRIQE-----KEEFENTRKNQR-----ALDSMQASLEAEAKGA 2057
QY 660 KTHNSYAL-ADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSLTURLQTQTQFRDMTTA- 717

Search completed: August 28, 2005, 10:49:18
Job time : 99 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:48:37 ; Search time 169 Seconds
(without alignments)

2343.607 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 5080

Sequence: 1 MKHLKTVALLTIVSVVTH.....LGYTSVALLSLTAKKKY 1008

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1767149 seqs, 392926209 residues

Total number of hits satisfying chosen parameters: 1767149

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 110 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	5080	100.0	1008	US-10-078-531-2
2	5076	99.9	1008	US-10-078-531-2
3	4901	96.5	971	US-10-078-531-7
4	4876	96.0	971	US-10-078-531-6
5	4871	95.9	969	US-10-078-531-8
6	4867	95.8	970	US-10-078-531-4
7	4831	95.1	963	US-10-078-531-5
8	4775	94.0	951	US-10-078-531-3
9	3777.5	74.4	1055	US-10-091-007-24
10	3777.5	74.4	1055	US-10-078-614-2
11	241.5	4.8	669	US-10-369-493-11351
12	11352	A		Sequence 11352, A
13	2948	Ap		Sequence 2948, Ap
14	254	App		Sequence 254, App
15	410	App		Sequence 410, App
16	4474	Ap		Sequence 4474, Ap
17	7463	A		Sequence 7463, A
18	62	Appl		Sequence 62, Appl
19	7646	Ap		Sequence 7646, Ap
20	135452			Sequence 135452
21	86	Appl		Sequence 86, Appl
22	70177	A		Sequence 70177, A
23	5834	Ap		Sequence 5834, Ap
24	81	Appl		Sequence 81, Appl
25	12996	A		Sequence 12996, A
26	440	App		Sequence 440, App
27	70580	A		Sequence 70580, A
28	23	Appl		Sequence 23, Appl
29	4098	Ap		Sequence 4098, Ap
30	43762	A		Sequence 43762, A
31	43811	A		Sequence 43811, A
32	5703	Ap		Sequence 5703, Ap
33	12611	A		Sequence 12611, A
34	321	App		Sequence 321, App
35	2	Appl		Sequence 2, Appl
36	2	Appl		Sequence 2, Appl
37	1586	Ap		Sequence 1586, Ap
38	15350	Ap		Sequence 15350, Ap
39	5635	Ap		Sequence 5635, Ap
40	12285	A		Sequence 12285, A
41	3334	Ap		Sequence 3334, Ap
42	3335	Ap		Sequence 3335, Ap
43	6	Appl		Sequence 6, Appl
44	196981	A		Sequence 196981, A
45	5942	Ap		Sequence 5942, Ap
46	11501	A		Sequence 11501, A
47	12913	A		Sequence 12913, A
48	5835	App		Sequence 5835, App
49	3302	Ap		Sequence 3302, Ap
50	862	App		Sequence 862, App
51	4927	Ap		Sequence 4927, Ap
52	99	Appl		Sequence 99, Appl
53	71235	A		Sequence 71235, A
54	5415	Ap		Sequence 5415, Ap
55	2070	Ap		Sequence 2070, Ap
56	3328	Ap		Sequence 3328, Ap
57	70176	A		Sequence 70176, A
58	3332	Ap		Sequence 3332, Ap
59	60608	A		Sequence 60608, A
60	5816	Ap		Sequence 5816, Ap
61	12967	A		Sequence 12967, A
62	220	App		Sequence 220, App
63	234	App		Sequence 234, App
64	1950	Ap		Sequence 1950, Ap
65	23124	A		Sequence 23124, A
66	46664	A		Sequence 46664, A
67	5639	Ap		Sequence 5639, Ap
68	11	Appl		Sequence 11, Appl
69	4	Appl		Sequence 4, Appl
70	2	Appl		Sequence 2, Appl
71	5739	Ap		Sequence 5739, Ap
72	100	App		Sequence 100, App
73	1839	Ap		Sequence 1839, Ap

```
85 185 3.6 3917 14 US-10-171-311-8 Sequence 8, Appli
86 185 3.6 3925 14 US-10-171-311-6 Sequence 6, Appli
87 185 3.6 5795 9 US-09-815-242-13610 Sequence 12610, A
88 184.5 3.6 1479 16 US-10-437-963-106122 Sequence 106122,
89 184 3.6 891 15 US-10-369-493-13477 Sequence 13477, A
90 184 3.6 1665 15 US-10-282-122A-71690 Sequence 71690, A
91 183 3.6 921 17 US-10-732-923-3305 Sequence 3305, Ap
92 183 3.6 1091 15 US-10-369-493-6328 Sequence 6328, Ap
93 183 3.6 1192 15 US-10-282-122A-57178 Sequence 57178, A
94 183 3.6 1440 16 US-10-437-963-168983 Sequence 168983,
95 182.5 3.6 944 9 US-09-815-242-11391 Sequence 11391, A
96 182.5 3.6 944 15 US-10-282-122A-58864 Sequence 58864, A
97 182 3.6 1192 9 US-09-815-242-10903 Sequence 10903, A
98 181.5 3.6 1786 9 US-09-742-096-3 Sequence 3, Appli
99 181.5 3.6 1787 15 US-10-415-253-2 Sequence 2, Appli
100 181.5 3.6 2026 9 US-09-801-368-86 Sequence 86, Appli
101 181.5 3.6 2026 15 US-10-369-493-22175 Sequence 22175, A
102 181.5 3.6 5171 16 US-10-408-765A-2687 Sequence 2687, Ap
103 180.5 3.6 1202 9 US-09-864-761-43061 Sequence 43061, A
104 180.5 3.6 1221 15 US-10-335-977-7813 Sequence 7813, Ap
105 180.5 3.6 1992 15 US-10-369-493-6527 Sequence 6527, Ap
106 178.5 3.5 995 9 US-09-864-761-49017 Sequence 49017, A
107 178 3.5 1212 9 US-09-815-242-5279 Sequence 5279, Ap
108 178 3.5 1224 18 US-10-617-320-4347 Sequence 4347, Ap
109 178 3.5 1482 9 US-09-815-242-12484 Sequence 12484, A
110 177.5 3.5 2076 9 US-09-815-242-5815 Sequence 5815, Ap
```

ALIGNMENTS

```
RESULT 1
US-10-078-531-2
; Sequence 2, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; TITLE OF INVENTION: DNA FRAGMENTS
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1008
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-2

Query Match 100.0%; Score 5080; DB 14; Length 1008;
Best Local Similarity 100.0%; Pred. No. 1e-303;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKHLKVALTTLTVSVVTHNQEVPSLVKEPIILKQTQASSISGADYAESGSKLKLINE 60
DB 1 MKKHLKVALTTLTVSVVTHNQEVPSLVKEPIILKQTQASSISGADYAESGSKLKLINE 60

QY 61 TSGPVDVDTVDFSDKRTTPKIKDNLAGPREQELKAVTENTESEKOITSGSQLEQSK 120
DB 61 TSGPVDVDTVDFSDKRTTPKIKDNLAGPREQELKAVTENTESEKOITSGSQLEQSK 120

QY 121 SLSLNTVPSTSNWEICDFTKGNLTVLGSLGSGVEKLSQTHLVLPQAAOGTQLIQVAS 180
DB 121 SLSLNTVPSTSNWEICDFTKGNLTVLGSLGSGVEKLSQTHLVLPQAAOGTQLIQVAS 180

QY 181 FAFTPDKKTAETYSRAGEGEISQDVGKGEIINEGEVFNLSYLLKKVTIPTGKHLGQ 240
```

```
Db 181 FAFTPDKKTAETYSRAGEGEISQDVGKGEIINEGEVFNLSYLLKKVTIPTGKHLGQ 240
QY 241 DAFVNDKNIAEAVNLPESETISDYAFALHALKQIDLPDNLKAIAGELAFDFOQITGKLSLP 300
DB 241 DAFVNDKNIAEAVNLPESETISDYAFALHALKQIDLPDNLKAIAGELAFDFOQITGKLSLP 300
QY 301 RQLMLRAERAPKSNHIIKTIEPRGNSLVKIVIGASQDNDLSQMLPDGKLEKIESAFTGNP 360
DB 301 RQLMLRAERAPKSNHIIKTIEPRGNSLVKIVIGASQDNDLSQMLPDGKLEKIESAFTGNP 360
QY 361 GDDHYNRVVLWTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKMLEEDFTYQKNSVT 420
DB 361 GDDHYNRVVLWTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKMLEEDFTYQKNSVT 420
QY 421 GFSNKGLOKVKRNKNLEIPKQHGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480
DB 421 GFSNKGLOKVKRNKNLEIPKQHGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480
QY 481 KIGAFAPQSNNLKSFASDDLEEKEGAFMNNRIETLELKDQKLVITIGDAAPHINIIAIV 540
DB 481 KIGAFAPQSNNLKSFASDDLEEKEGAFMNNRIETLELKDQKLVITIGDAAPHINIIAIV 540
QY 541 LPESVQETIGRSAPFRQNGANNLI FMGSVKVTIGEMAFISNRLEHLDLSEQKOLTEIPVOAF 600
DB 541 LPESVQETIGRSAPFRQNGANNLI FMGSVKVTIGEMAFISNRLEHLDLSEQKOLTEIPVOAF 600
QY 601 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSIAFNALDDDDGDFQFNKVVVK 660
DB 601 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSIAFNALDDDDGDFQFNKVVVK 660
QY 661 THNSYALADGEHPIVDPDKLSSTIVDLKILKILIEGLDYSTLTQTTQTQPRDMTTAGKA 720
DB 661 THNSYALADGEHPIVDPDKLSSTIVDLKILKILIEGLDYSTLTQTTQTQPRDMTTAGKA 720
QY 721 LLSKSNLRQGEKQKPLQEAOPFLGRVDLDKAIKAEKALVTKKATKQGLLERSINKAVL 780
DB 721 LLSKSNLRQGEKQKPLQEAOPFLGRVDLDKAIKAEKALVTKKATKQGLLERSINKAVL 780
QY 781 AYNSAIKCANVKRLEKELDLTLGLVEGKGPLAQATWVQGVYLLKTPLPPEYIIGLVNY 840
DB 781 AYNSAIKCANVKRLEKELDLTLGLVEGKGPLAQATWVQGVYLLKTPLPPEYIIGLVNY 840
QY 841 PDKSGKLIYALDMSDTTIGEGOKDAYGNPILNVDNEDNGYHALAVATLADYEGLDIKTILN 900
DB 841 PDKSGKLIYALDMSDTTIGEGOKDAYGNPILNVDNEDNGYHALAVATLADYEGLDIKTILN 900
QY 901 SKLSQLTSSIRQVPTAAVHRAGI FOAIONAAAEAEQQLPKPQTHSEKSSSSSANSKORGL 960
DB 901 SKLSQLTSSIRQVPTAAVHRAGI FOAIONAAAEAEQQLPKPQTHSEKSSSSSANSKORGL 960
QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLITAIKKKKY 1008
DB 961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLITAIKKKKY 1008
```

```
RESULT 2
US-10-474-792-46
; Sequence 46, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Wintter, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 46
; LENGTH: 1008
```

; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
; US-10-474-792-46.

Query Match 99.9%; Score 5076; DB 16; Length 1008;
Best Local Similarity 99.9%; Pred. No. 1.8e-303;
Matches 1007; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKGHLKVALTLTVSVVTHNQVFSLVKPEILKQTQASSISGADYABSSGSKLKINE 60
DB 1 MKGHLKVALTLTVSVVTHNQVFSLVKPEILKQTQASSISGADYABSSGSKLKINE 60

QY 61 TSGPVDVDTDLFSDKRTTPEKIKONLAKGPRQELKAVTENTPESKQITSGSQBSKE 120
DB 61 TSGPVDVDTDLFSDKRTTPEKIKONLAKGPRQELKAVTENTPESKQITSGSQBSKE 120

QY 121 SLSLNKTVSTSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQADGTQLIQVAS 180
DB 121 SLSLNKTVSTSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQADGTQLIQVAS 180

QY 181 FAFTPDKKTAAIAYTSRAGENGESIQDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
DB 181 FAFTPDKKTAAIAYTSRAGENGESIQDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240

QY 241 DAFVDMKNIAEVLNLPESLETISDYAPAHALAKQIDLPDNLKAIGELAFDNOITGKLSLP 300
DB 241 DAFVDMKNIAEVLNLPESLETISDYAPAHALAKQIDLPDNLKAIGELAFDNOITGKLSLP 300

QY 301 ROLMRLAERAFKSNHKTTFPRGNSLKVI GEASFOQNDLSQMLPDGLEKISEAPTGNP 360
DB 301 ROLMRLAERAFKSNHKTTFPRGNSLKVI GEASFOQNDLSQMLPDGLEKISEAPTGNP 360

QY 361 GDDHYNRNVVLTKSGKPNPGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB 361 GDDHYNRNVVLTKSGKPNPGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

QY 421 GFSNKGLOKVRKNLLEIPKOHNGVTITIGNNAFNVDFONKTLRKYLDEEVKLPSTIR 480
DB 421 GFSNKGLOKVRKNLLEIPKOHNGVTITIGNNAFNVDFONKTLRKYLDEEVKLPSTIR 480

QY 481 KIGAFAPQSNLKSFEASDDLEIIEKGAFFMNRRIETLEKDKLVITGDAAFHNIHIAIV 540
DB 481 KIGAFAPQSNLKSFEASDDLEIIEKGAFFMNRRIETLEKDKLVITGDAAFHNIHIAIV 540

QY 541 LPESVQEIERSAFRONGANLIFMGSKVKTGEMAFSLNRLHLDLSEKQITFIPVQAF 600
DB 541 LPESVQEIERSAFRONGANLIFMGSKVKTGEMAFSLNRLHLDLSEKQITFIPVQAF 600

QY 601 SDNALKEVLLPASLKTIREEAFKQNLKOLEVASALSHIAFNALDDNDGDEQFONKVVVK 660
DB 601 SDNALKEVLLPASLKTIREEAFKQNLKOLEVASALSHIAFNALDDNDGDEQFONKVVVK 660

QY 661 THNSVALADGHEFIVDPDKLSSTIVDLKILKLEGLDYSTRQTQTQFRDMTTAGKA 720
DB 661 THNSVALADGHEFIVDPDKLSSTIVDLKILKLEGLDYSTRQTQTQFRDMTTAGKA 720

QY 721 LLSKSNLRQGEKQKLOEAFPLGRVLDLKAIAKAEKALVTKATNGOLLERSINKAVL 780
DB 721 LLSKSNLRQGEKQKLOEAFPLGRVLDLKAIAKAEKALVTKATNGOLLERSINKAVL 780

QY 781 AYNNSAIKANVKRLEKELDLTLGLVEGKPLAQATWQGVLLKTPPLPEYIYGLNVI 840
DB 781 AYNNSAIKANVKRLEKELDLTLGLVEGKPLAQATWQGVLLKTPPLPEYIYGLNVI 840

QY 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKTIIN 900
DB 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKTIIN 900

QY 901 SKLSQITSIQVPTAAHYHRAFIQAIQNAABABOLLPKPGTHSEKSSSESANSKDRGL 960
DB 901 SKLSQITSIQVPTAAHYHRAFIQAIQNAABABOLLPKPGTHSEKSSSESANSKDRGL 960

QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYVILGYTSVALLSIIITAIKKKKY 1008

DB 961 QSNPKTNRGRHSAILPRTGSKGSFVYVILGYTSVALLSIIITAIKKKKY 1008

RESULT 3
US-10-078-531-7
; Sequence 7, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; TITLE OF INVENTION: DNA FRAGMENTS
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-7

Query Match 96.5%; Score 4901; DB 14; Length 971;
Best Local Similarity 100.0%; Pred. No. 9.9e-293;
Matches 971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LVKEPILKQTQASSISGADYABSSGSKLKINETSGPVDVDTDLFSDKRTTPEKIKON 86
DB 1 LVKEPILKQTQASSISGADYABSSGSKLKINETSGPVDVDTDLFSDKRTTPEKIKON 60

QY 87 LAKGPREQLKAVTENTSEKQITSGSQBSKEKSLNKTVPSTSNWEICDFTKGNL 146
DB 61 LAKGPREQLKAVTENTSEKQITSGSQBSKEKSLNKTVPSTSNWEICDFTKGNL 120

QY 147 VGLSKSGVEKLSQTDHLVLPQADGTQLIQVASFAFTPDKKTAAIAYTSRAGENGESIQ 206
DB 121 VGLSKSGVEKLSQTDHLVLPQADGTQLIQVASFAFTPDKKTAAIAYTSRAGENGESIQ 180

QY 207 LDVDCKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDMKNIAEVLNLPESLETISDYAF 266
DB 181 LDVDCKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDMKNIAEVLNLPESLETISDYAF 240

QY 267 AHLALKQIDLPDNLKAIGELAFDNOITGKLSQMLPDGLEKISEAPTGNP 326
DB 241 AHLALKQIDLPDNLKAIGELAFDNOITGKLSQMLPDGLEKISEAPTGNP 300

QY 327 KVI GEASFOQNDLSQMLPDGLEKISEAPTGNPDHYNRNVLTWTKSGKPNPGLATEN 386
DB 301 KVI GEASFOQNDLSQMLPDGLEKISEAPTGNPDHYNRNVLTWTKSGKPNPGLATEN 360

QY 387 TVNVPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNLLEIPKOHNGVT 446
DB 361 TVNVPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNLLEIPKOHNGVT 420

QY 447 ITEIGDNAPRVDFONKTLRKYLDEEVKLPSTIRKIGAFAPQSNLKSFEASDDLEIIEKE 506
DB 421 ITEIGDNAPRVDFONKTLRKYLDEEVKLPSTIRKIGAFAPQSNLKSFEASDDLEIIEKE 480

QY 507 GAFWNNRLETLEKDKLVITGDAAFHNIHIAIVLPESVQEIERSAFRONGANLIFMGS 566
DB 481 GAFWNNRLETLEKDKLVITGDAAFHNIHIAIVLPESVQEIERSAFRONGANLIFMGS 540

QY 567 KVKTLGEMAFSLNRLHLDLSEKQITFIPVQAFSDNALKEVLLPASLKTIREEAFKQNH 626
DB 541 KVKTLGEMAFSLNRLHLDLSEKQITFIPVQAFSDNALKEVLLPASLKTIREEAFKQNH 600

```
Qy 627 LKQLEVASALSHIAFNALDDNDGDQFQFNKVVVKTTHNSYALADGEHFI VDPDKLSSTIV 686
Db 601 LKQLEVASALSHIAFNALDDNDGDQFQFNKVVVKTTHNSYALADGEHFI VDPDKLSSTIV 660
Qy 687 DLEKILKILIEGLDYSTLRQTQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRV 746
Db 661 DLEKILKILIEGLDYSTLRQTQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRV 720
Qy 747 DLDKAIKAEKALVTKKATNGQQLLERSINKAVLAYNNSAIKKANVKELEKELDLTLGLV 806
Db 721 DLDKAIKAEKALVTKKATNGQQLLERSINKAVLAYNNSAIKKANVKELEKELDLTLGLV 780
Qy 807 EGKGLAQATWVGQVYLLKTPLEPYIYGLNVYFDKSGKLIYALDMSDTTIGEGQKDAYG 866
Db 781 EGKGLAQATWVGQVYLLKTPLEPYIYGLNVYFDKSGKLIYALDMSDTTIGEGQKDAYG 840
Qy 867 NPILNVDEDEGEGYHALAVATLADYEGLDIKTILNSKLSQTSIRQVPTAAVHRAGIFQAI 926
Db 841 NPILNVDEDEGEGYHALAVATLADYEGLDIKTILNSKLSQTSIRQVPTAAVHRAGIFQAI 900
Qy 927 QNAAAEAEQLLPKPGTHSEKSSSESANSKDRGLOSNEKPTRGRHSATLPRTGSGSFVY 986
Db 901 QNAAAEAEQLLPKPGTHSEKSSSESANSKDRGLOSNEKPTRGRHSATLPRTGSGSFVY 960
Qy 987 GILGYTSVALL 997
Db 961 GILGYTSVALL 971
```

RESULT 4

```
US-10-078-531-6
; Sequence 6, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 971
; TYPE: PRP
; ORGANISM: Streptococcus pyogenes
US-10-078-531-6
```

```
Query Match 96.0%; Score 4876; DB 14; Length 971;
Best Local Similarity 99.4%; Pred. No. 3.4e-291;
Matches 965; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 27 LVKEPILKQTOASSISGADYAESGKGLKINETSGPVDDTVTDLPFSDKRTTPEKIKDN 86
Db 1 LVKEPILKQTOASSISGADYAESGKGLKINETSGPVDDTVTDLPFSDKRTTPEKIKDN 60
Qy 87 LAKGPRBELKAVTENTSEKQITSGSLQESLSLNKTVPSNWEICDFTIKGNTL 146
Db 61 LAKGPREQLKTVTENTSEKQITSGSLQESLSLNKTVPSNWEICDFTIKGNTL 120
Qy 147 VGLSKSGVEKLSQTHLVLPQAAAGTQLIQVAFPTPKKTAIAYTSRAGENGEISQ 206
Db 121 VGLSKSGVEKLSQTHLVLPQAAAGTQLIQVAFPTPKKTAIAYTSRAGENGEISQ 180
Qy 207 LDVDGKEIINEGEVFNLSYLLKKVTIPTGYKHKIQDAFVDNKNIAEAVNLPSLETISDYAF 266
Db 181 LDVDGKEIINEGEVFNLSYLLKKVTIPTGYKHKIQDAFVDNKNIAEAVNLPSLETISDYAF 240
```

```
Qy 267 AHLALKOIDLPDNLKALGELAFPDNQITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSL 326
Db 241 AHLALKOIDLPDNLKALGELAFPDNQITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSL 300
Qy 327 KVIAGEASFQDNDLSQLMPLDGLKIESEAFITGNPGDDHNNRVVLTMTKSGKNPGLATEN 386
Db 301 KVIAGEASFQDNDLSQLMPLDGLKIESEAFITGNPGDDHNNRVVLTMTKSGKNPGLATEN 360
Qy 387 TYNVPDKSLWOESPEIDYTKWLEBDFTYQKNSVTGFSNKGLOKVRKNKNIIEIPKHNGVT 446
Db 361 TYNVPDKSLWOESPEIDYTKWLEBDFTYQKNSVTGFSNKGLOKVRKNKNIIEIPKHNGVT 420
Qy 447 ITEIGDANFRNVDQNTLRKYDLEEVKLPSTIRKIGAFQFQSNLNKSFASDDLEETKE 506
Db 421 ITEIGDANFRNVDQNTLRKYDLEEVKLPSTIRKIGAFQFQSNLNKSFASDDLEETKE 480
Qy 507 GAFMNNRIETLEKDKLVTTIGDAAFHINHIYAIIVLPESVQIEGRSAFRQNGANNLIFMGS 566
Db 481 GAFMNNRIETLEKDKLVTTIGDAAFHINHIYAIIVLPESVQIEGRSAFRQNGANNLIFMGS 540
Qy 567 KVTILGEMAFILSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASIKTIREEAPKKNH 626
Db 541 KVTILGEMAFILSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASIKTIREEAPKKNH 600
Qy 627 LKQLEVASALSHIAFNALDDNDGDQFQFNKVVVKTTHNSYALADGEHFI VDPDKLSSTIV 686
Db 601 LKQLEVASALSHIAFNALDDNDGDQFQFNKVVVKTTHNSYALADGEHFI VDPDKLSSTIV 660
Qy 687 DLEKILKILIEGLDYSTLRQTQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRV 746
Db 661 DLEKILKILIEGLDYSTLRQTQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRV 720
Qy 747 DLDKAIKAEKALVTKKATNGQQLLERSINKAVLAYNNSAIKKANVKELEKELDLTLGLV 806
Db 721 DLDKAIKAEKALVTKKATNGQQLLERSINKAVLAYNNSAIKKANVKELEKELDLTLGLV 780
Qy 807 EGKGLAQATWVGQVYLLKTPLEPYIYGLNVYFDKSGKLIYALDMSDTTIGEGQKDAYG 866
Db 781 EGKGLAQATWVGQVYLLKTPLEPYIYGLNVYFDKSGKLIYALDMSDTTIGEGQKDAYG 840
Qy 867 NPILNVDEDEGEGYHALAVATLADYEGLDIKTILNSKLSQTSIRQVPTAAVHRAGIFQAI 926
Db 841 NPILNVDEDEGEGYHALAVATLADYEGLDIKTILNSKLSQTSIRQVPTAAVHRAGIFQAI 900
Qy 927 QNAAAEAEQLLPKPGTHSEKSSSESANSKDRGLOSNEKPTRGRHSATLPRTGSGSFVY 986
Db 901 QNAAAEAEQLLPKPGTHSEKSSSESANSKDRGLOSNEKPTRGRHSATLPRTGSGSFVY 960
Qy 987 GILGYTSVALL 997
Db 961 GILGYTSVALL 971
```

RESULT 5

```
US-10-078-531-8
; Sequence 8, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 8
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-8

Query Match      95.9%; Score 4871; DB 14; Length 969;
Best Local Similarity 99.6%; Pred. No. 7e-291;
Matches 965; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 28 VKPEILKQTOQASSISGADYAESSGSKLKNETSGPVDVDTVTLFSDKRTTPEKIKDNL 87
DB 1 VKPEILKQTOQASSISGADYAESSGSKLKNETSGPVDVDTVTLFSDKRTTPEKIKDNL 60
QY 88 AKGPREQELKAVTENTSEKQITSGSQESKSLNKTVPSTSNWEICDFITKGNLTIV 147
DB 61 AKGPREQELKAVTENTSEKQINSQESKSLNKTVPSTSNWEICDFITKGNLTIV 120
QY 148 GLSKSGVEKLSQTDHLVLPQAAQTQLIQVASFAPTPDKKTAIAEYTSRAGENGESQL 207
DB 121 GLSKSGVEKLSQTDHLVLPQAAQTQLIQVASFAPTPDKKTAIAEYTSRAGENGESQL 180
QY 208 DVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPSLETSIDYAPA 267
DB 181 DVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPSLETSIDYAPA 240
QY 268 HLALKQIDLPDNLKALGELAFDNOITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSLK 327
DB 241 HLALKQIDLPDNLKALGELAFDNOITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSLK 300
QY 328 VTGEASFQDNDLSQMLPDGLEKIESEAFNGPGDDHNNRVLVMTKSGNPSGLATENT 387
DB 301 VTGEASFQDNDLSQMLPDGLEKIESEAFNGPGDDHNNRVLVMTKSGNPSGLATENT 360
QY 388 YNPNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPIKQHNGVTI 447
DB 361 YNPNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPIKQHNGVTI 420
QY 448 TEIGNAFNRVDFQNKTLRKYLEEVKLPSTIRKIGAPAFQSNLKSFEASDDLEIKEG 507
DB 421 TEIGNAFNRVDFQNKTLRKYLEEVKLPSTIRKIGAPAFQSNLKSFEASDDLEIKEG 480
QY 508 AFMNNRIETLEKDKLVTTIGDAAFINHIYAIVLPESVQIEGRSAFRQNGANNLIIFMGSK 567
DB 481 AFMNNRIETLEKDKLVTTIGDAAFINHIYAIVLPESVQIEGRSAFRQNGANNLIIFMGSK 540
QY 568 VKTLGEMAPLSNRLEHLDLSEQKQLTEIPVQAFSONALKEVLLPASLKTIRBEAFKKNHL 627
DB 541 VKTLGEMAPLSNRLEHLDLSEQKQLTEIPVQAFSONALKEVLLPASLKTIRBEAFKKNHL 600
QY 628 KQEVASALSHIAFNALDNDGDQFDNKKVVKTHNSVALADGHEFIVDPDKLSTTIVD 687
DB 601 KQEVASALSHIAFNALDNDGDQFDNKKVVKTHNSVALADGHEFIVDPDKLSTTIVD 660
QY 688 LEKILKLEGLDYSTLRQTTQTFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVD 747
DB 661 LEKILKLEGLDYSTLRQTTQTFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVD 720
QY 748 LDKAIAKAEKALVTKATQNGQLLERSINKAVLAYNNSAIKKNVVRLEKELDLTLGLVE 807
DB 721 LDKAIAKAEKALVTKATQNGQLLERSINKAVLAYNNSAIKKNVVRLEKELDLTLGLVE 780
QY 808 GKQPLAQATWQGVLLKTPLEPYIIGLVNYPFKSGKLIYALDMSDTIIGEGQDAYGN 867
DB 781 GKQPLAQATWQGVLLKTPLEPYIIGLVNYPFKSGKLIYALDMSDTIIGEGQDAYGN 840
QY 868 PILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQLTSLRQVPTAAYHRAGIFQAIQ 927
DB 841 PILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQLTSLRQVPTAAYHRAGIFQAIQ 900
QY 928 NAAAEAEQLLPKPGTHSEKSSSESANSKORGLQNPKNRGRHSAILPRTGSKGSFVYG 987
DB 901 NAAAEAEQLLPKPGTHSEKSSSESANSKORGLQNPKNRGRHSAILPRTGSKGSFVYG 960
```

```
QY 988 ILGYTSVAL 996
DB 961 ILGYTSVAL 969

RESULT 6
US-10-078-531-4
; Sequence 4, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 970
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-4

Query Match      95.8%; Score 4867; DB 14; Length 970;
Best Local Similarity 99.5%; Pred. No. 1.2e-290;
Matches 965; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 27 LVKPEILKQTOQASSISGADYAESSGSKLKNETSGPVDVDTVTLFSDKRTTPEKIKDNL 86
DB 1 LVKPEILKQTOQASSISGADYAESSGSKLKNETSGPVDVDTVTLFSDKRTTPEKIKDNL 60
QY 87 LAGPREQELKAVTENTSEKQITSGSQESKSLNKTVPSTSNWEICDFITKGNLTIV 146
DB 61 LAGPREQELKAVTENTSEKQINSQESKSLNKTVPSTSNWEICDFITKGNLTIV 120
QY 147 VGLSKSGVEKLSQTDHLVLPQAAQTQLIQVASFAPTPDKKTAIAEYTSRAGENGESQL 206
DB 121 VGLSKSGVEKLSQTDHLVLPQAAQTQLIQVASFAPTPDKKTAIAEYTSRAGENGESQL 180
QY 207 LDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPSLETSIDYAP 266
DB 181 LDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPSLETSIDYAP 240
QY 267 AHLALKQIDLPDNLKALGELAFDNOITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSL 326
DB 241 AHLALKQIDLPDNLKALGELAFDNOITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSL 300
QY 327 KVIAGEASFQDNDLSQMLPDGLEKIESEAFNGPGDDHNNRVLVMTKSGNPSGLATEN 386
DB 301 KVIAGEASFQDNDLSQMLPDGLEKIESEAFNGPGDDHNNRVLVMTKSGNPSGLATEN 360
QY 387 TVNPNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPIKQHNGVT 446
DB 361 TVNPNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPIKQHNGVT 420
QY 447 ITEIGDNAPNRVDFQNKTLRKYLEEVKLPSTIRKIGAPAFQSNLKSFEASDDLEIKKE 506
DB 421 ITEIGDNAPNRVDFQNKTLRKYLEEVKLPSTIRKIGAPAFQSNLKSFEASDDLEIKKE 480
QY 507 GAFMNNRIETLEKDKLVTTIGDAAFINHIYAIVLPESVQIEGRSAFRQNGANNLIIFMGSK 566
DB 481 GAFMNNRIETLEKDKLVTTIGDAAFINHIYAIVLPESVQIEGRSAFRQNGANNLIIFMGSK 540
QY 567 KVKTILGEMAPLSNRLEHLDLSEQKQLTEIPVQAFSDNALKEVLLPASLKTIRBEAFKKNH 626
```

Db 541 KVKTLGEMAFSLNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREBAFKQNH 600
Qy 627 LKQLEVASALSHIAFNALDDNDGDEQDFNKKVVKTHNSYALADGEHFIVDPDKLSSTIV 686
Db 601 LKQLEVASALSHIAFNALDDNDGDEQDFNKKVVKTHNSYALADGEHFIVDPDKLSSTIV 660
Qy 687 DLEKTLKILIEGLDYSTLQTTQTOTQFRDMMTAGKALLSKNLRQGEKQFLQBAQFPLGRV 746
Db 661 DLEKTLKILIEGLDYSTLQTTQTOTQFRDMMTAGKALLSKNLRQGEKQFLQBAQFPLGRV 720
Qy 747 DLDKAIAKAEKALVTKKATKQQLLERSINKAVLAYNNSAIKCANVVKLEKELDLTLGLV 806
Db 721 DLDKAIAKAEKALVTKKATKQQLLERSINKAVLAYNNSAIKCANVVKLEKELDLTLGLV 780
Qy 807 EGKGLQAQTMVQGYVLLKTPLEPYVIGLVNVPDKSGKLIYALDMSDTTIGEGQKDAYG 866
Db 781 EGKGLQAQTMVQGYVLLKTPLEPYVIGLVNVPDKSGKLIYALDMSDTTIGEGQKDAYG 840
Qy 867 NPILNVDENEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVHRAGIFQAI 926
Db 841 NPILNVDENEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVHRAGIFQAI 900
Qy 927 QNAAAEASQLLPKPGTHSEKSSSESANSKORGLOSNPKTNRGRHSAILPRTGSKGSFVY 986
Db 901 QNAAAEASQLLPKPGTHSEKSSSESANSKORGLOSNPKTNRGRHSAILPRTGSKGSFVY 960
Qy 987 GILGYTSVAL 996
Db 961 GILGYTSVAL 970

RESULT 7

US-10-078-531-5
; Sequence 5, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 963
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-5

Query Match 95.1%; Score 4831; DB 14; Length 963;

Best Local Similarity 99.4%; Pred. No. 2e-288;

Matches 957; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 34 KOTQASSISGADYAESGSKLKTNETSGPVDVTDVTLFSDKRTTPBKIKDNLAKGPKE 93
Db 1 KOTQASSISGADYAESGSKLKTNETSGPVDVTDVTLFSDKRTTPBKIKDNLAKGPKE 60
Qy 94 QELKAVTENTESEKQITSGSQLEQSKESLSLNKTVPSTSNWEICDFITKGNLTVLGSKSG 153
Db 61 QELKAVTENTESEKQINGSGSQLEQSKESLSLNKRVPSNWEICDFITKGNLTVLGSKSG 120
Qy 154 VEKLSQTDHLVLPQAAQDTQLIQVAFPTPKKTAIAEYTSRAGEGEISQLDVDGKE 213
Db 121 VEKLSQTDHLVLPQAAQDTQLIQVAFPTPKKTAIAEYTSRAGEGEISQLDVDGKE 180
Qy 214 IINEGEVFNLSLLKKVTIPTGYKHIGQDAFVNDKNKNAEVLNPESLETISDYAFALHAKQ 273

Db 181 IINEGEVFNLSLLKKVTIPTGYKHIGQDAFVNDKNKNAEVLNPESLETISDYAFALHAKQ 240
Qy 274 IDLPDNLKAIGELAFFNQITGKLSPLQRLMRLAERAFKSNHIKTIFRGNLSKVIQGEAS 333
Db 241 IDLPDNLKAIGELAFFNQITGKLSPLQRLMRLAERAFKSNHIKTIFRGNLSKVIQGEAS 300
Qy 334 FQDNDLSQMLPDGLEKIESEAFPTGNDDHYNRVVLTWSKGNPSGLATENTYVNPDK 393
Db 301 FQDNDLSQMLPDGLEKIESEAFPTGNDDHYNRVVLTWSKGNPSGLATENTYVNPDK 360
Qy 394 SLWQESPEIDYTWLWLEBEDFTYQKNSVTGFSNKGLOKVKRNKNLEIPKOHNGVTITTEIGDN 453
Db 361 SLWQESPEIDYTWLWLEBEDFTYQKNSVTGFSNKGLOKVKRNKNLEIPKOHNGVTITTEIGDN 420
Qy 454 AFRNVDFONKTLRKYLEEVKLPSTIRKIGAFAPQSNLKSFEASDDLEIEKEGAFMNNR 513
Db 421 AFRNVDFONKTLRKYLEEVKLPSTIRKIGAFAPQSNLKSFEASDDLEIEKEGAFMNNR 480
Qy 514 IETLELKKDLVTIGDAAPHNHIIYAVLPRESVQIEGRSAFRONCANNLIIPGSKVKTLGE 573
Db 481 IETLELKKDLVTIGDAAPHNHIIYAVLPRESVQIEGRSAFRONCANNLIIPGSKVKTLGE 540
Qy 574 MAPLSNLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREBAFKQHLKQLEVA 633
Db 541 MAPLSNLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREBAFKQHLKQLEVA 600
Qy 634 SALSHIAPNALDDNDGDEQDFNKKVVKTHNSYALADGEHFIVDPDKLSSTIVLEKILK 693
Db 601 SALSHIAPNALDDNDGDEQDFNKKVVKTHNSYALADGEHFIVDPDKLSSTIVLEKILK 660
Qy 694 LIEGLDYSTLRQTOTQFRDMMTAGKALLSKNLRQGEKQFLQBAQFPLGRVLDKAIA 753
Db 661 LIEGLDYSTLRQTOTQFRDMMTAGKALLSKNLRQGEKQFLQBAQFPLGRVLDKAIA 720
Qy 754 KAERKALVTKKATKQQLLERSINKAVLAYNNSAIKCANVVKLEKELDLTLGLVEGKGPLA 813
Db 721 KAERKALVTKKATKQQLLERSINKAVLAYNNSAIKCANVVKLEKELDLTLGLVEGKGPLA 780
Qy 814 QATVQGVYLLKTPLEPYVIGLVNVPDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVD 873
Db 781 QATVQGVYLLKTPLEPYVIGLVNVPDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVD 840
Qy 874 EDNEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVHRAGIFQAIQNAAAEA 933
Db 841 EDNEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVHRAGIFQAIQNAAAEA 900
Qy 934 EQLLPKPGTHSEKSSSESANSKORGLOSNPKTNRGRHSAILPRTGSKGSFVYGILGYTS 993
Db 901 EQLLPKPGTHSEKSSSESANSKORGLOSNPKTNRGRHSAILPRTGSKGSFVYGILGYTS 960
Qy 994 VAL 996
Db 961 VAL 963

RESULT 8

US-10-078-531-3
; Sequence 3, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 951
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-3

Query Match 94.0%; Score 4775; DB 14; Length 951;
Best Local Similarity 99.3%; Pred. No. 5.5e-285;
Matches 944; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 46 DYAESGSKLKINETSQVDDTVDLFSDKRTTPEKIKONLAKGPREQELKAVTENTES 105
DB 1 DYAESGSKLKINETSQVDDTVDLFSDKRTTPEKIKONLAKGPREQELKAVTENTES 60

QY 106 EKQITSGSQLEQSKESLSLNKVPSTSNWEICDFITKGNLTVLGSKSGVEKLSQTDHLVL 165
DB 61 EKQITSGSQLEQSKESLSLNKVPSTSNWEICDFITKGNLTVLGSKSGVEKLSQTDHLVL 120

QY 166 PQQAADGTOLIOVAFPAFTPKDKTAIAEYTSRAGENGEISQDVGKEIINEGEVFNLSYL 225
DB 121 PQQAADGTOLIOVAFPAFTPKDKTAIAEYTSRAGENGEISQDVGKEIINEGEVFNLSYL 180

QY 226 LKKVTIPTGYKHIGQDAFVDNKNIAEVNLPESLETISDYAFALHALKQIDLDPNLKAIGE 285
DB 181 LKKVTIPTGYKHIGQDAFVDNKNIAEVNLPESLETISDYAFALHALKQIDLDPNLKAIGE 240

QY 286 LAFPDNQITGKLSLPROLMLRAERAFKSNHIIKTIEPRGNSLKVIGEASFQDNDLSQMLP 345
DB 241 LAFPDNQITGKLSLPROLMLRAERAFKSNHIIKTIEPRGNSLKVIGEASFQDNDLSQMLP 300

QY 346 DGLKIESEAFNGPDDHNNRVLVLTSGKNPGLATENTYVNPDKSLWQESPEIDYT 405
DB 301 DGLKIESEAFNGPDDHNNRVLVLTSGKNPGLATENTYVNPDKSLWQESPEIDYT 360

QY 406 KWLDEEFTYQKNSVTGFSNGKIQKVRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTL 465
DB 361 KWLDEEFTYQKNSVTGFSNGKIQKVRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTL 420

QY 466 RYVDLEEVKLPSTIRKIGAFQSNLKSFEASDDLEETKEGAFMNNRIETLEKDKLVT 525
DB 421 RYVDLEEVKLPSTIRKIGAFQSNLKSFEASDDLEETKEGAFMNNRIETLEKDKLVT 480

QY 526 IGDAAFHINHIIYAI VLPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLD 585
DB 481 IGDAAFHINHIIYAI VLPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLD 540

QY 586 LSEQQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSIAFNALD 645
DB 541 LSEQQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSIAFNALD 600

QY 646 DNDGDEQFNKVVKTHNSYALADGEHPI VDPDKLSSTIVLEKILKILIEGLDYSTRQ 705
DB 601 DNDGDEQFNKVVKTHNSYALADGEHPI VDPDKLSSTIVLEKILKILIEGLDYSTRQ 660

QY 706 TTQTFQPRDMMTAGKALLSKSNLROCKOKFLOEAFPLGRVDLDKAIKAERKALYTKKAT 765
DB 661 TTQTFQPRDMMTAGKALLSKSNLROCKOKFLOEAFPLGRVDLDKAIKAERKALYTKKAT 720

QY 766 KNGQLLERSINKAVLAYNNSAIKKANVRKLEKELDOLLTGLVGEKGPLAQATVWQGVYLLK 825
DB 721 KNGQLLERSINKAVLAYNNSAIKKANVRKLEKELDOLLTGLVGEKGPLAQATVWQGVYLLK 780

QY 826 TPLPLPEYIIGLVNVPDKSGKLIYALDMSDTTGEQOKDAYGNPILNVDEBNDEGYHALAVA 885
DB 781 TPLPLPEYIIGLVNVPDKSGKLIYALDMSDTTGEQOKDAYGNPILNVDEBNDEGYHALAVA 840

QY 886 TLADVEGLDKITLNSKLSQTSIROVPTAAVHRAGI FOAIONAAEAEBOLLPKPGTHSE 945
DB 841 TLADVEGLDKITLNSKLSQTSIROVPTAAVHRAGI FOAIONAAEAEBOLLPKPGTHSE 900

QY 946 KSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVAL 996

DB 901 KSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVAL 951

RESULT 9
US-10-091-007-24
; Sequence 24, Application US/10091007
; Publication No. US20030170782A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics limited
; APPLICANT: Le Page, Richard W F Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21978WO
; CURRENT APPLICATION NUMBER: US/10/091,007
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: GB 9921125.2
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-091-007-24

Query Match 74.4%; Score 3777.5; DB 14; Length 1055;
Best Local Similarity 71.7%; Pred. No. 1.5e-223;
Matches 756; Conservative 105; Mismatches 140; Indels 53; Gaps 5;

QY 2 KKHKTVALTLTTVSVVTHNQEVSLVKEPILKQTPASSISGADYABESSGSKSLKINET 61
DB 3 KKHKTVALTLTTVSVVTHNQEVSLVKEPILKQTPASSISGADYABESSGSKSLKINET 61

QY 62 SGPVDDTVDLFSDKRTTPEKIKONLAKGPREQELKAVTENT-ESEKQITSGSQLEQSK 120
DB 62 NSTVDETVSDLFDGNSNNSSKTESVVSDFKQVPEKAPTEVTQEASNSNDASKVEVPKQ 121

QY 121 SLSLNKTVPTSNWEICDFITKGNLTVLGSKSGVEKLSQTDHLVLPSSAADGTOLIOVAS 180
DB 122 DTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPSSAADGTOLIOVAS 181

QY 181 FAFTPDKKTAAEYTSRAGENGEISQDVGKEIINEGEVFNLSYLLKVVITPTGYKHIGQ 240
DB 182 FAFTPDKKTAAEYTSRAGENGEISQDVGKEIINEGEVFNLSYLLKVVITPTGYKHIGQ 241

QY 241 DAFVONKNIAEVNLPESLETISDYAFALHALKQIDLDPNLKAIGELAFPDNQITGKLSLP 300
DB 242 DAFVONKNIAEVNLPESLETISDYAFALHALKQIDLDPNLKAIGELAFPDNQITGKLSLP 301

QY 301 RQMLAERAFKSNHIIKTIEPRGNSLKVIGEASFQDNDLSQMLPDLGLEKIESEAFNGP 360
DB 302 RHLIKLAERAFKSNRIQTVEFLGSKLVIGEASFQDNNLRNVLDPGLEKIESEAFNGP 361

QY 361 GDDYHNNRVLVLTSGKNPGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB 362 GDEYHNNQVVLTRTGQNPQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVT 421

QY 421 GFSNKGLOKVRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480
DB 422 GFSNKGLOKVRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 481

QY 481 KIGAFQSNLKSFEASDDLEETKEGAFMNNRIETLEKDKLVTIGDAAPHINHIIYAI V 540
DB 482 KIGAFQSNLKSFEASDDLEETKEGAFMNNRIETLEKDKLVTIGDAAPHINHIIYAI V 541

QY 541 LPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSEOKQLTEIPVQAF 600
DB 542 LPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSEOKQLTEIPVQAF 601

QY 601 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSIAFNALDNDGDEQFNKVVVK 660
DB 602 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSIAFNALDNDGDEQFNKVVVK 661

```
QY 661 THNSYALADGHRFIVDPDKLSTIVDLKILKLEGLDYSTLRTOTQOTQFRDWTAGKA 720
DB 662 THNSHMLADGHRFIVDPDKLSTIVDLKILKLEGLDYSTLRTOTQOTQFRDWTAGKA 721
QY 721 LLSKNLRQGEKQKFLQBAQFPLGRVDLDKAKAEKALVTKATKNGOLLERSINKAVL 780
DB 722 LLSKNLRQGEKQKFLQBAQFPLGRVDLDKAKAEKALVTKATKNGOLLERSINKAVL 781
QY 781 AYNSAIKANKVRLKLELDLTLGLVEGKPLAQATMVGQVYLLKTPPLPEYIYGLNVY 840
DB 782 AYNSAIKANKVRLKLELDLTLGLVEGKPLAQATMVGQVYLLKTPPLPEYIYGLNVY 841
QY 841 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAVATLADYEGLDIKTILN 900
DB 842 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAVATLADYEGLDIKTILN 901
QY 901 SKLSOLTSTIROVPTAAVHRAGIFQAIQNAAAEQLLPKPGTHSEKSSSSANSKDRGL 960
DB 902 SBLDKIKAIRQIPLAKYHRLGIFQAIRNAAEADRLPK-----TPKGYLNEVPNTRKKQM 957
QY 961 QSNPK-----TN-----RGRHSA 973
DB 958 EKNLKPVDYKTPFNKALPNEKVDGDRAAKGHNINAEITNNSVAVTPIRSEQQLHKSQSDV 1017
QY 974 ILPRTSGKSFVYIGILYTSVALLSLITAIKXXX 1007
DB 1018 NLPQTSSKNNFYIILGYVSLCLLFLVTAGKKGK 1051

RESULT 10
US-10-476-614-2
; Sequence 2, Application US/10476614
; Publication No. US2004017113A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem Inc.
; TITLE OF INVENTION: Antigens of Group B Streptococcus and Corresponding DNA Fragments
; FILE REFERENCE: 74872-81
; CURRENT APPLICATION NUMBER: US/10/476,614
; CURRENT FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: 2003-11-03
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Streptococcus
US-10-476-614-2

Query Match 74.4%; Score 3777.5; DB 16; Length 1055;
Best Local Similarity 71.7%; Pred. No. 1.5e-223;
Matches 756; Conservative 105; Mismatches 140; Indels 53; Gaps 5;

QY 2 KKHLLKVALTLTVSIVTHNQVPSIVKPEILKOTQASSISGADYABSSGSKLKNET 61
DB 3 KKHLLKVALTLTVSIVTHNQVPSIVKPEILKOTQASSISGADYABSSGSKLKNET 61
QY 62 SGVDDVTDLPSDKRTTPEIKDNLAKGRPEQLKAVTENT-ESEKQITSGSQLEQSK 120
DB 62 NSTVETVSDLSFDGNSNNSSKTESVSDPKQVPKAPKPEVTOEASNSNDASKVEVPKQ 121
QY 121 SUSLNTVPSTNWEICDPITKNTLVGLSKGVBLKSGTDHLVLPQAAADGTQLIQVAS 180
DB 122 DPASKKETLETSTWEAKDPVTRGDTLVGFSKSGINKLSQTSHLVLPQAAADGTQLTQVAS 181
QY 181 FAPTPDKKTAIAEYTSRAGENCEISOLDVDGKEIINEGEVENSLLKVKVTPTGKHLGQ 240
DB 182 FAPTPDKKTAIAEYTSRAGENCEISOLDVDGKEIINEGEVENSLLKVKVTPTGKHLGQ 241
QY 241 DAFVDMKNAEVLNPSLETISDYAFALHAKQIDLPNLKAIAGELAFPDNQITKGLSLP 300
DB 242 DAFVDMKNAEVLNPSLETISDYAFALHAKQIDLPNLKAIAGELAFPDNQITKGLSLP 301
```

```
QY 301 RQLMRLAERAPKSNHIIKTIIEPRGNSLKVIGEASPDQNDLSQMLPDGLEKIESEAFQNP 360
DB 302 RHLIKLAERAPKSNRIQTVBFLGSKLVIGEASPDQNNLRNMLPDGLEKIESEAFQNP 361
QY 361 GDDHYNRVVLWTSGKNPSGLATENTVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB 362 GDEHYNQVVLRTKQGNPHQLATENTVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 421
QY 421 GFSNGKQKQVKNRNKLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480
DB 422 GFSNGKQKQVKNRNKLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 481
QY 481 KIGAFAPQSNMLKSFASDDLEEBKEGAFPMNRITETLEKOKLVITIGDAAPIHIIYAIV 540
DB 482 KIGAFAPQSNMLKSFASDDLEEBKEGAFPMNRITETLEKOKLVITIGDAAPIHIIYAIV 541
QY 541 LPESVQETIGRSAPFQNGANNLI PMGSVKVTIGEMAFLSNRLEHLDLSQKOLTIPVQAF 600
DB 542 LPESVQETIGRSAPFQNGANNLI PMGSVKVTIGEMAFLSNRLEHLDLSQKOLTIPVQAF 601
QY 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSASHIAFNALDDNDGDQFQNKVVVK 660
DB 602 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSASHIAFNALDDNDGDQFQNKVVVK 661
QY 661 THNSYALADGHRFIVDPDKLSTIVDLKILKLEGLDYSTLRTOTQOTQFRDWTAGKA 720
DB 662 THNSHMLADGHRFIVDPDKLSTIVDLKILKLEGLDYSTLRTOTQOTQFRDWTAGKA 721
QY 721 LLSKNLRQGEKQKFLQBAQFPLGRVDLDKAKAEKALVTKATKNGOLLERSINKAVL 780
DB 722 LLSKNLRQGEKQKFLQBAQFPLGRVDLDKAKAEKALVTKATKNGOLLERSINKAVL 781
QY 781 AYNSAIKANKVRLKLELDLTLGLVEGKPLAQATMVGQVYLLKTPPLPEYIYGLNVY 840
DB 782 AYNSAIKANKVRLKLELDLTLGLVEGKPLAQATMVGQVYLLKTPPLPEYIYGLNVY 841
QY 841 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAVATLADYEGLDIKTILN 900
DB 842 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAVATLADYEGLDIKTILN 901
QY 901 SKLSOLTSTIROVPTAAVHRAGIFQAIQNAAAEQLLPKPGTHSEKSSSSANSKDRGL 960
DB 902 SBLDKIKAIRQIPLAKYHRLGIFQAIRNAAEADRLPK-----TPKGYLNEVPNTRKKQM 957
QY 961 QSNPK-----TN-----RGRHSA 973
DB 958 EKNLKPVDYKTPFNKALPNEKVDGDRAAKGHNINAEITNNSVAVTPIRSEQQLHKSQSDV 1017
QY 974 ILPRTSGKSFVYIGILYTSVALLSLITAIKXXX 1007
DB 1018 NLPQTSSKNNFYIILGYVSLCLLFLVTAGKKGK 1051
```

```
RESULT 11
US-10-369-493-11351
; Sequence 11351, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11351
```

```
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(669)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-11351

Query Match
Best Local Similarity 4.8%; Score 241.5; DB 15; Length 669;
Matches 124; Conservative 84; Mismatches 214; Indels 125; Gaps 28;

QY 136 ICDFITKGNLTGVLGSKGVKLSQTDHVLPSQAADGTOLIQVAFPTPKKTAIAEYT 195
Db 177 LCMSTTHGAATVSSSEQTLPXLRLT--LTVGLTSDGTS-VTVTGVY-GPDSDVIP--- 229

QY 196 SRAGENGEISQDVGKEIINEGEVFNYSYL--LKKVTPTGKYHIGQDAFVNDKNAEYN 253
Db 230 -----GEI-----DGLPVTTIGASACKALTSLTSPDSTVTTIDASAFYGTNLSVT 278

QY 254 LPSELETSDYAF-AHLALKQIDLDPNKAIGELAFEDNQITGKLS--LPRQLMELAEAA 310
Db 279 MPDSVTITIGSSAFYGTGLASVTIIPNSVTITIGSSAFYGTGLISVTIPNSVTITIGSSA 336

QY 311 FKS-NHIKTIIEPRGNSLKVIGE-----ASFQDNDLSQLMLPDGLEKIES 353
Db 337 FYGCTNLVTMMENGNAPTTVGSNWASGNTLVAYISEGATNFTTPWNEVPCYPALTAATA 396

QY 354 E-----AFTG---NPGDDHYNNRVVLWTKGKNSP--GLATENTY----- 388
Db 397 EFTSNIVYGASPLTVKFTYTGVCANALDWYFNDSTV--DSTARNPSTYTDPTGYSIKLN 455

QY 389 --VNPDKSLWQESPEID--YTKWLE--EDFTYQKN-----SVTGFSGKGLQKVENKLEIP 439
Db 456 VSNP-----WGSDEKVTAYITVAEPVDNFTYLSGDTSTVITGYSQPG-----GDVVIP 504

QY 440 KOHNGVTITEIGDGNAPRVDFQNKTLRKVDLEVKLPSTIRKIGAFAPQS--NNLKSFEAS 498
Db 505 STIGNLPVTAISDSVPKS-----WTNITSVIPDSVTITIGSSAFYGTGLTSVIP 555

QY 499 DDLEIKGAFNN-NRIETLEKDKLVITIGDAAFH--INHIIAIVLPESVQEIIGRSAPRQN 556
Db 556 NSVTITIGNAFNGCTGLTSMITGNVTKIDRMFYGCSALISWIPDSVTITIGDSAFYQ- 614

QY 557 GANNLIFMGSKVKTLGEMAFNLRLHLDLSEKQQLTEIPVOAFSD-NALKEVLLPASLK 615
Db 615 -CSNL-----ASVTIPDSVTITIDSYAFRDCNALTSVTIPDSVT 651

QY 616 TIREBAF 622
Db 652 NVESRLP 658
```

```
RESULT 12
US-10-369-493-11352
; Sequence 11352, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11352
```

```
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
US-10-369-493-11352

Query Match
Best Local Similarity 4.7%; Score 237.5; DB 15; Length 668;
Matches 128; Conservative 94; Mismatches 207; Indels 211; Gaps 20;

QY 13 TTVSVVTHNQEVFSIAVKEPILKQTAQAS-----SSISGADYAEBSCKSKLKNETSGPVDD 67
Db 147 TSDIGVLYNKDVDTLIRCPGKTSVPAIPNSVTITISSVFPDSCGLTITIGSGVTSIDN 206

QY 68 TTVDLFSKRTTPKIKONLAKGPREQELKAVTENTESEKQITSGSQLESKESLSLNTK 127
Db 207 SV-----FSDC-----SALTAL-----DVDNDNSMYASIDGVLYNK- 237

QY 128 VPSTSNWEICDPITKGNLTGVLGSKGVKLSQTDHVLPSQAADGTOLIQVAFPTPK 187
Db 238 -----DITTLIQCP-----DITTLIQCP----- 247

QY 188 KTAIAEYTSRAGENGEISQDVGKEIINEGEVFNYSYLKKVTIPTGKYHIGQDAFVNDK 247
Db 248 -----GKTGSVT--IPDSVTITIGDSAFYGTALTSTVITIPDSVTITIGTSAPQGC 294

QY 248 NIAEVLNPESLETSDYAFAH-LALKQIDLDPNKAIGELAFEDNQITGKLSLPRQLMRL 306
Db 295 SITSNIGNNVTSIGNSAFCDICALTSTVIGNNVISIGSAFQGCALTSVTIPDSVTKI 354

QY 307 AERAPKSNHIKTIIEPRGNSLKVIGESAF-QDNDLSQLMLPD----- 346
Db 355 GSSAFQGCISITSVNVISGVTSGNSAFQCSALTFTVTIPDSVTITILDYAFQDCSRLTTV 414

QY 347 ----GLEKISEAFTGNPDGDDHYNNRVVLWTKGKNSPGLATENTYVNPDKSLWQESPEI 402
Db 415 TIGSVTNIDSAFSG-----CSALTAID--VDNDNSIYASIDGV 452

QY 403 DYTQWLEEDFTYQKNSVTGFSNKGKQVKRNKNLEIPKQHNGVTITEIGDGNAPRVDD--- 459
Db 453 LYNK-----DITTLIQCPG-----KTSVITIP-----SVTIGDSAFRDCSGLT 493

QY 460 -----FQNKTLRKVDLEVKLPSTIRKIGAFAPQS--NNLKSFEASDDLEE 503
Db 494 IVPIGNNVTSIGSSAFYGTCTV---LTSVPIGNNVTSIGSSAFYGTCTSLTSVTIGNNVTS 549

QY 504 IKEGAFNN-NRIETLEKDKLVITIGDAAFH-----INNIIYA 538
Db 550 ISSAFRQCSALTFTVTIPDSVTITIGDIAPDSCGLTITVIGNNVTSIGSSAFYGTALT 609

QY 539 IVPESVQEIIGRSAPRQNGANNLIFMGSKVKTLGEMAFLS 578
Db 610 IAPDSVTISIGSAPRDCNTLTSVTIGNNVTSIGSSAFRS 649
```

```
RESULT 13
US-10-617-320-2948
; Sequence 2948, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
```


Db 648 HODELERLSQLEL-----ELNFKHAQBIYEKNLDEKAKEISNLNQLIEEFKQAD 699
Qy 559 NN-----LIFMGSKVKTIGENAFISNLEHLDL-----SEQQLTEIPQAFS 601
Db 700 NNSAFTALSERDQLL-----SQVELSMVTEJRAQVKQLEMNLAERQRLDYSESQTAH 756
Qy 602 DNALKEVL-----LPASLKTIRBEAFKQNLKQLEVASALSHIAFNALDNDGDEQDNK- 656
Db 757 DNLLTEQIHSLSIEAKSKDVKIEVL-QNELDDVQLQFSEQSTLIRSL-----QSOLQNK 810
Qy 657 -----VVVKTTHNSVALADGEHPIVDPKLSSTIVDLKILKIEGLDYSTLR 704
Db 811 SEVLEGAERVRHISKVELSQALSQKELEITMDQLL-----LEK-----KRDVETIQ 859
Qy 705 QTTOTQFRDMMTAGKALLSKSNLRQCEKQKQFLOEAFILGRVLDLQAKAEKALVTKKA 764
Db 860 QYIEKQDQVT-----EISFMTKVMQVNEEKFSLG-VEI-KTLKEQLNLLSRAEE 909
Qy 765 TNGQLLERSINKAVLAYNNSAIKKA---NVKRLKEKELDLLTGLVEGKGLAQATWVGQV 821
Db 910 AKKEQVEEDNEVSGLKQNYDEMSAPAGQISKEELQHEFDLLAKKENEQRKRKLQAALINRK 969
Qy 822 YLLKTPPLPPEYIYGLNVFDSKGLIYALDMSDT-IGEGQKDAYGNPILNVDEDEGHH 880
Db 970 ELLQVSRLEBELANLK---DESKK---EIPLSETERGEVEE-----DKNKEYS 1013
Qy 881 ALAVATLADYEGDLDTILNLSQLTSIR-----QVPTAAVHRAGI 922
Db 1014 EKCVTSKQOEIYLKQITISEKEVELQHIRKOLEKLAABEQFQALVKOMNTOLOKTNQ 1073
Qy 923 FOAIQNAABAEQLPKPQTHSEKSSSES-ANSKDRGLQSNPKTNRGH 971
Db 1074 IDLLQAEISENQAIQKLTISNTDASDGSVALVKETVVISPCGSSSH 1123

RESULT 15
US-10-474-776-410
; Sequence 410, Application US/10474776
; Publication No. US20040110181A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING POLYE
; FILE REFERENCE: AM100649-PCT
; CURRENT APPLICATION NUMBER: US/10/474,776
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 752
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 410
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-474-776-410

Query Match 4.2%; Score 213.5; DB 16; Length 621;
Best Local Similarity 23.2%; Pred. No. 0.00014;
Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31;
Qy 6 KTVALTTLTVSVV-----THNQEVF-----SLVKEPILK-----Q 35
Db 3 KTTILSLTTAAVILAAYVNEPIADTPSEVIKTKVGSIIQNNIKYKVLTVEGNIGT 62
Qy 36 TQASSISGADY-ABSSGK-----SKLKINETSQVDDTVTDLFS-----DK 76
Db 63 VQVNGVTPVEFEAGQDGKPTPTKITYGDKVFTVTEVASQAFSYYPDETGRIVYPPSS 122
Qy 77 RTTPEKIKDLAKGPREQELKAVTENTESEKQITSGSQLEQSK----- 120
Db 123 ITIPSSIKKIQQKGFHGSKAKTII-----FDKGSQLEKIEDRAFDFSELEELPAS 174
Qy 121 -----SLSLNK-TVPSTSNWEICDFTITKGNLTVGLSKSGVEKLSQTDHLVPSQ 168

Db 175 LEYIGTSAFSFSQKLKKLTFSSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222
Qy 169 AAD-GTQLIOVA-----SPA-----FTPDKTAIAEYTSRAGENGESISOLD 208
Db 223 VKTIGSNLFRUTTSKTVVDVEEGNESFASVGVLFPSKD-KTQLIYIPSQ--KNDESYPKP 279
Qy 209 VDGEIINEGEVFNFSYLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPSLETISDYAP-A 267
Db 280 KETRELASYSFNKNSY-LKLELNEGLEKIGTFAPADAIKLEELSLPNSLETIERLAFVG 338
Qy 268 HIALKQIDLPDLNKAIGELAF-----FDNQITKLSLPROLMRLAERAFKSNHIK 317
Db 339 NLELKEILPDPNVKPNFGKRVNGLPKLSLTIGNNIN---SLSPFSLSGVLDLSLKEIHK 395
Qy 318 -----TIEFRGNS-LKVIGEASFQ-DNDLSQLMLPDGLEKIESEAPT 357
Db 396 NKSTEPSVKKDTFPAIPEVTKFVTVSEHIKDVLSKNLSTSNDI-----IVEKVDNIKQETDV 451
Qy 358 GNPQDDHYNNRVVLWTK-----SGKNPSGLATENTYVNPDKSLW---QESPEIDY 404
Db 452 AKP-KKNSNQGVGVKDKGLWYLYNESGSMATGWV-----KDKGLWYLYNESGSM-A 502
Qy 405 TKWLEED--FTYQKNS---VTGF-SNKGL 427
Db 503 TGWVKDKGLWYLYNESGSMATGWVKDKGL 531

RESULT 16
US-10-472-928-4474
; Sequence 4474, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; PRIOR FILING DATE: 2003-09-26
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWinn9, version 1.03
; SEQ ID NO 4474
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: choline binding protein PcpA (pcpA)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Feature of note: WY motif
; OTHER INFORMATION: Similar to strain R6 sequence 15903986 (0.B+01)
US-10-472-928-4474

Query Match 4.2%; Score 213.5; DB 17; Length 621;
Best Local Similarity 23.2%; Pred. No. 0.00014;
Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31;
Qy 6 KTVALTTLTVSVV-----THNQEVF-----SLVKEPILK-----Q 35
Db 3 KTTILSLTTAAVILAAYVNEPIADTPSEVIKTKVGSIIQNNIKYKVLTVEGNIGT 62
Qy 36 TQASSISGADY-ABSSGK-----SKLKINETSQVDDTVTDLFS-----DK 76
Db 63 VQVNGVTPVEFEAGQDGKPTPTKITYGDKVFTVTEVASQAFSYYPDETGRIVYPPSS 122
Qy 77 RTTPEKIKDLAKGPREQELKAVTENTESEKQITSGSQLEQSK----- 120
Db 123 ITIPSSIKKIQQKGFHGSKAKTII-----FDKGSQLEKIEDRAFDFSELEELPAS 174
Qy 121 -----SLSLNK-TVPSTSNWEICDFTITKGNLTVGLSKSGVEKLSQTDHLVPSQ 168
Db 175 LEYIGTSAFSFSQKLKKLTFSSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222


```
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7646
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Candida albicans
; NAME/KEY: MISC FEATURE
; LOCATION: (1881)..(1881)
; OTHER INFORMATION: X-any amino acid
US-10-032-585-7646

Query Match      4.2%; Score 211; DB 14; Length 1881;
Best Local Similarity 19.9%; Pred. No. 0.00094;
Matches 227; Conservative 173; Mismatches 431; Indels 310; Gaps 48;

QY 5 LKTVLTLTVSVVTHN-----QEVSLVKEPILKQTOASSIS 43
Db 689 LPTVLTVDVYVNLVDNFIRIKRALFHPDPSPRGKISYELFEELDTKVLNLTKEIQ--T 746

QY 44 GADYAESGSKLKNINETSQVDDVTDLFSDKRTTPEKIDNLAGPREQELKAVTENT 103
Db 747 EKENAESNDK---ELNEKI-----EKLNLSTKLET---KLED-----KEQELAKIQEDH 790

QY 104 ES--EKQITSGQLSQSKESLSINKTVPSTNSWEICDFTTKNTLVGLSKGVEKLSQTD 161
Db 791 KSLNEKFLVTANSLCGIKARTKESITSGPDQQLQELQKALKKGNTSESTLQKLKEDSTGE 850

QY 162 -----HLVPSQAADGTOLIQ-----VASPAFTPDKTAIAEYTSR 197
Db 851 QAKKKLEDGINNMTDRDLFHLKXSKSEAE--TQIKQREPERPNLTYSFENTKDYELQINN 909

QY 198 AGENGEISO-----LDVDGK-----EIIINEGEVFNLLKKVTIPTGVKKHG 239
Db 910 KNSNNEFPKQINELSKKTESLTDNKNFAKQLEKLRDTEENNEHMDKLRAS-----VA 965

QY 240 QDAFVDNKNIAE---VNLPSLETISDYAFALHAKQIDLPNL--KAIGELAFFNQITG 295
Db 966 YNDLAKAKSESEETVKAKELETLSKI-----DNLEKELKQSKKKELEG 1013

QY 296 KLSLPRLQMLRAERAPK--SNIHKITEPGNSLKVIGEASFDNDLSQLMLPDGLEKIES 353
Db 1014 QL---QNTDSTNEXPFKELEDELSIK-----KSNKEISSQNSBLIQ-----KLEKTEK 1059

QY 354 EAPTGNPGDHYNNRVNLTKSGKNPSGLATENTVYVDPKSLWQSPEDIDYTKWLEEDFT 413
Db 1060 DL---QAKDEIDKLKAEFK-----SNIDNLSNLSISSQSKLKEA-----EESH 1101

QY 414 YQKNSVTGFSNKGLOKVK---RNKNLETPKHQNGVTITEIGDNAPRVNDFQNKTLRYD 469
Db 1102 STKDEHSSLS--ENLKKLKEVENTKTSMAK-----LSAKIEBHKKATDEIETKTKHTD 1155

QY 470 LEEVKLPSTIRKIGAFAPSNLKS-----PEASDLEEI---KEGAPMNRLETLEL 519
Db 1156 LQE---EHAQKQSFESERNDIKNLDEANKELSDNREKLSNLEKEKTELNKKLTOEE 1211

QY 520 KDKLVTIGDAAPHINHIYAIVLPESVQEIGRSAPRQNGANNLIFMGSKVKTIGEMAFLSN 579
Db 1212 K---ISDLETSVAISEDKSKSLKHIEDLKR-----EKIKL---ETTLKEN 1251

QY 580 RLEHLDLSQKQJTEIPVQAFSDNALKEVLLPASLKTIIE-----EAPKKNH 626
Db 1252 EETMPEKKEQLQVW-----NDCKE--LEACLKLTETKEKEINLIRKLEAKSDH 1301

QY 627 LKQLEVASALSHAFNALDNDGDGFQFNKVVVVKTHNSYAL-----668
Db 1302 DTERKKLSL-----IDTKSESE---KNVKNLEQTEKLGEBREKEVRDIQSLAAK 1351

QY 669 -ADGBHFVDPDKLSSTIVDLBKILKLEGLDYSTLRQTTQT-----QFRDMWTTA 717
```

```
Db 1352 TTDWEKIKTTLDKVLKEKSDLEKTNK--ESVD--TLKKEVENLAKKEISLLEDQKDDTTK 1407
QY 718 GKALLSKSNLRQGBKQKFLQEAQFFLGRVDLDKAIKAEKALVTYKKA TK---NQOLLERS 774
Db 1408 YKELAAQLE-----TKTSNLDSTTMELEKTELELKKVRNELTEATSELTQLDNNQSLTEE 1463
QY 775 INKAVLAYNNS-----AIK--KANVKRL-----EKELD 800
Db 1464 LEKTKAALTSSKDLVCGNQSLEQDLSKSVKSELKNFNENKYNQETSLSKDEIEEKQKE 1523
QY 801 LLTGLVBEKGPLAQATWVQGVYLLKTPLPPLPEYVYIGLVNVPYDKGKLIYALDMSDTTIGEG 860
Db 1524 IVTLQTELKORISEVEKERAMLSENSETVIVEKSDKISLESKINSIKENHSKITTHNE 1583
QY 861 QKDAYGNPILNVDEDNEG-----YHALAVATLADYEGLOIKITLN 900
Db 1584 QKTSLKQDIAKLSQDHSQAQTLQEDKENQLKELKASLEKHNKTESATSTEENKNOIKELSE 1643
QY 901 SKLSQLSIRQVPTAAVHRAGIFQAIQNAARAEASQLLPKPGTHSEK--SSSESANSKDRG 959
Db 1644 TIKSLTELKTSGDALKQSQKYEKTKLTKNSDTSKLEKQLEBELEKVKSDLQTADEKLDG 1703
QY 960 L 960
Db 1704 I 1704

RESULT 20
US-10-437-963-135452
; Sequence 135452, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbasuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 135452
; LENGTH: 2295
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37127C.1.pep
US-10-437-963-135452

Query Match      4.1%; Score 210; DB 16; Length 2295;
Best Local Similarity 20.1%; Pred. No. 0.0014;
Matches 215; Conservative 176; Mismatches 411; Indels 270; Gaps 47;

QY 16 SVWTHNQEVPSLVKEPILKQTOASSISGADYAESGSKSKLNINETSQVDDTVTDLPD 75
Db 950 ALLVAEQEENDLTAKAHANAQERNEELS--KEVEDADGKIK-----QLSDTVQRL--- 997

QY 76 KRTTPEKTIKNLAGPREQELKAVTENTESKQITSGQLSQSKESLS--LNKTVPTSTSNW 134
Db 998 BETIQEREAALLAERQEKESASAVIABSQARNE-AFASKLEDAEKQIDLLQETV---QRF 1053
QY 135 EICDPITKGNLTVLGSKSGVEKLSQTDHLVPSQA--ADGTOLIOVASPAFTPDKKTATAE 193
Db 1054 E--BAITKQLQSV-----TIEKQHEETVQLAAQAQAKIDELLREA---GDTDEKSTQLET 1104

QY 194 YTSRAGENETISQLDV-----DGKEIINEGEVFNLSYLLKKVVTIPTGVKKIGQDAF 243
Db 1105 TIQRLLES--LTENDALLTTERQETATKLSLSAQYKNBELLKKI-----EDA-- 1151
```



```

Db      1336  TEKENLVIANSKLNELHONKLELOVALAAVAEKEGSEETHSLRKTLDGMIOKE 1395
QY      706   TTQOTFRDMTTAGKALLSKSNURQGEKQKFLQRAQFFLGRVDLDKAIKAEKALVTKKAT 765
Db      1396  ELESQVSSVTEHHEELSKSYNTLSEKQ--LLNEKYESAKKELGEAIAKLEBQMNVDKSE 1453
QY      766   KNGQL--LERSINKAVLAYNNNAIKKANVRLKELDLTLGLVEGKPLAQATWQGYVL 823
Db      1454  KELHISKLERQITLSELKTYMBE--IOTMQVETTEKDEALTTRKOE-----HANLVH---- 1502
QY      824   LKTPLEPLPEYIIGLVNFDKSKGLIYALDMSDTIGEGOKDAYGNGPILNVDED--NEGHAL 882
Db      1503  -----EKOALEQQLEVRKELNDAYH-- 1523
QY      893   AVATLADVEGLDIKILNSKLSQLSIRQVPTAAVHR--AGIFQAIQNAAEAE 934
Db      1524  ---TIANOE-----EQSVREIKWDAYKKSSEDQLEAEQQHVABE 1561

```

RESULT 22

```

US-10-282-122A-70177
; Sequence 70177, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangou
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```

```

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70177
; LENGTH: 3533
; TYPE: PRT
; ORGANISM: Staphylococcus aureus

```

```

US-10-282-122A-70177

```

```

Query Match      4.1%; Score 207; DB 15; Length 3533;
Best Local Similarity 19.2%; Pred. No. 0.004;
Matches 231; Conservative 169; Mismatches 424; Indels 380; Gaps 52;

```

```

QY      18   VTHNQEVSLVKEPILKQTQASSISGADYAESSGSKL-----KINETSG----- 63
Db      1515  VONAESSIAGLNNPNTINK-----GNVSSATQAVISSKNALDQGVRLAQDQKQAGNSLNHL 1570
QY      64   ---PVDDTVTDLDFDKRTPEKIKDLAKGPREGE--LKAVENTESEKQITSGSQL----- 115
Db      1571  QLTFAQQALENQINNATTRDKVAEIIAQALNEMAKLESIKDQPTQESKFINED 1630
QY      116  BOSKESISL-----LNKTVPTSTNWEICDFTIKGNTLVGLSKSGVEKLSQTDHLVLP 166
Db      1631  QAQKDAVTAQVQAHAKDLINKTTDPTLAKSIIDQATQAVTDAKNLHGDQKLAQ-----D 1684
QY      167  SOAADGTQLIOVASFATPPDKKTAIAEYTSAGENGELSOLDVDGKEIINEGEVFN 226
Db      1685  KORA--TETLNNLSNLNTPQEQ--ALENQINNAATRGEVAQKLTQAALNQAMEALRSIQ 1741
QY      227  KKVITPTGYKHIGQD-----AFVNKN--IAEVNLP-----ESLETISDYAFAPHL 269
Db      1742  DQQTEAGSKFINEDKPKDAYQAQVQNAKDLINQTNPTLDKQAVEQLTOAVNQAKDNL 1801
QY      270  ALKQIDLPDNLKAI GELAFPDNQITGKLSLPRQLMLAERAFKSNHIXITIFRGNLSKVI 329
Db      1802  HGDQKLAADDKQHAVTDL-----NQLNGLNPNPQAL-----ESQINNAATRGEVAQKL 1849
QY      330  GEASFQNDLSQL--MLPDGLEKTESAFTG--NPGDDHYNRVVLWTSGKNSFGSLATE 385
Db      1850  ABAKALQAMQALRNSIQDQQQTESGSKFINEDKPKDAYQAQV-----QNAKDLINQ 1902
QY      386  NTVNPDKSLMQESPEIDYTKWLBEDFTYQKNSVTGFSNKGLOKVKRKN----- 434
Db      1903  TGNPTLKSQVEQ-----LTQAVTTAKDNL-----HGDQKLARDQQQAVTTVNALP 1948
QY      435  NLEIPKQ-----HNGVTITEIGDNAPRVNVDQNKTLRKYDLBEVLKPSIRKIGAPAF 487
Db      1949  NLNHAQQQALTDAINAAPTRTEVAQH-----QTATELDHAMETLK--NKVDQVNTDKA 2000
QY      488  QSNMLKFEASDDLEBEKEGAP-----MNNRI 514
Db      2001  QPNYT---EASTDKKEAVDQALQAAESITDPTGNSNANKDAVDQVLTCLKSEKELNGNE 2057
QY      515  ETLEKDKLVTIGDAAAPHINHIYAILVPESVQETGRSAFRONGANNLIFMGSVKVTLCEM 574
Db      2058  RVAEAKTQAKQTIQDLTHLN-----ADQIATA--KQN-----IDQATKLOPIAEL 2100
QY      575  AFLSNRL-----EHLDSLSEKQLTEIPVQAFSD--NALKEVLLPASLKTIRREE 620
Db      2101  VDQATQLNQSMDQLQQA VNEHANVEQTVDYT---QADSDKQNAVQKAIADA-----EN 2150
QY      621  APKGNHLKQLEVASALSIAFNALDDNDGDGFQFNKVUVKTH--H-----NSYALADGEH 673
Db      2151  VLKQNAKQ--QVDQALQNI--LNKQALNGDSEV---ALAKTNGKHDIQDLNALNNAQODG 2205
QY      674  F-----IVDP-----DKLSSTIVDL----- 688
Db      2206  PKGRIDQSDNLNQIQIIVDEAKALNRAMDQLSQBITNEGRTKGSTNYVNAQTQVKQYVD 2265
QY      689  EKILKLEGLDYSTLRQTTQ---PRDMMTAGKALLSKSNLRQGEKQKFLQ----- 737
Db      2266  ETVDKAKQALDKSTGQNLTAQVILKNDVATAAKKALNGEBRLNRRKAEALQRLDQLTHL 2325
QY      738  ---EAQFPLGRVDLDKAIKAEKALVTKKATNGSQLL---ERSINKAVLA----- 781
Db      2326  NNAQRQLAIQQINNAETLNKASRAI--NRATKLDNAMGVSQOYIDRSHGLGVISSTNYINA 2383
QY      782  -----YNN-----SAIKKANVKELEKELDLTLGLVEGKPLAQATWQVQ 820
Db      2384  DDNLKANYDNAINAANAHELDKVGQNAKAEAEQKQNIIDAQNALNGDQNLNAK----- 2439
QY      821  VYLLKTPLEPPEYIIGLVNFPDKSKGLIYALDMSDTIGEGOKDAYGNPILNVDE----- 874
Db      2440  -----DKANAFVNSL---NGLNQOQOQDLAHKAINNADTVSDVTD 2475

```

```
QY 875 -----DNEGYYHALAVATLADYEGLD-----IKTILASK 902
Db 2476 IVNNQIDLDNAMEDTLKHLVDNEIPNA---EQVTNYQNADNNAKTNFDDAKRLANTLLNSD 2532
QY 903 LSQLTSIR---QVPTAAYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSESANSGKRG 959
Db 2533 NTNVDINGAIOAVDAIHNLNGDQRLQDAKDKATQISNOALANKLKEIEASNATDQDKL 2592
QY 960 LQSN 963
Db 2593 IAKN 2596

RESULT 23
US-10-470-048B-81
; Sequence 81, Application US/10470048B
; Publication No. US2005003744A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; FILE REFERENCE: SONN:035US
; CURRENT APPLICATION NUMBER: US/10/470,048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
; LENGTH: 1992
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1222)..(1225)
; OTHER INFORMATION: X = anything
US-10-470-048B-81

Query Match 4.1%; Score 206.5; DB 17; Length 1992;
Best Local Similarity 19.0%; Pred. No. 0.0019;
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;

QY 46 DYAESGSKKLINETSGBVDVTDLFSDKRTTPEKIKDNIAKPREOELKAVNTES 105
Db 19 DHAKQTVSQAHLNNAQKHMEDTLI-----DSETTRTAVKQDLTEAQAALDQALDQALDQQA 74
QY 106 EKOITSGSQLEQSKESLSLNTKVPSTSNWEICDFITKGNLTGLSKSGVEK---LSOTDH 162
Db 75 DKDATRASSAYNAEPNKKQSDVEAVQNAE-----SIIAGLNPTINKGNVSSATQA 126
QY 163 LVLPSQAADGTQLI-----QVASFA-----FTPDKKTAIAEYTSRAGENGESQLDQVQK 212
Db 127 VISSKNALDGVRLAQDKQTAGNSLNHLDLQTPAQQAALLENQINNATTRGEVAQKLETAQ 186
QY 213 EIINGEVNSVLLKKVTPTGYKHIGQD-----AFVDNKN---IAEVNLP-----255
Db 187 ALNQAMEALRNSIQDQOQTEAGSKFNEDEKPKDAYQAAVQNAKOLINTNNPTLDKQAV 246
QY 256 ESELTISDYAFAPHLAKQIDLPDLNKAIGELAFFNQITGKLSLPROMLRAERAFKSNH 315
Db 247 EQLTQAVNAQKNLHGDKQKLADQKHAVTDL-----NQLNGLNPNQOAL-----ESQ 294
QY 316 IKTIIFRGNLSKVIGEASFQDNDLSQ---MLPDGLEKIESEAFGTG--NFGDDHYNNRVVL 371
Db 295 INNAATRGEVAQKLAELAKALDQAMQALRNSIQDQOQTESGSKFNEDEKPKDAYQAAV-- 352
QY 372 WTKSGNPSGLATENTYNNPDKSLMQESPEIDYTKWLEEDFTYQKNSVTGFSNKGQKVK 431
Db 353 -----QNAKDLINQTNPTLDKQSVQEQ-----LQAVTTAKDNL-----HGPQKLA 393
QY 432 RNK-----NLEIPKQ-----HNGVTITEIGDNRNVDVFNKTLRKVDLEEV 473
Db 394 RDOQQAVTTVNALPNLNHQAQQALTDAINAAPTRTEVAQHV-----QTATELDHA METL 447
QY 474 KLPSTIRKIGAFQFQSNLKSPEASDLEIEKGAPE-----509
```

```
Db 448 K--NKVDQVNTDKAQPNYT---EASTDKKEAVDQALQAAESITDPTGNSNANKDAVDQVL 502
QY 510 -----MNNRIETLELKDKLVITIGDAAFHINHIYAIVLPESVQIEGRSAFRONGANN 560
Db 503 TKLOEKENELNGNERVAEAKTOAKQITDOLTHLN-----ADQIATA--KON-----546
QY 561 LIFGSKVKTIGEMAFLSNRL-----EHLDLSEQKQLTEIPVQAFPSD--NALK 606
Db 547 -IDQATKLOPIAELVDQATLQNSMDLQQAQVNEHANVEQTVDYT---QADSDKQNAK 601
QY 607 EVLLPASLKTTRIEAFKKNHLKQLEVASALSHIAFNALDDNDGDQFQFNKVVVKTTH--H- 663
Db 602 QAIADA-----ENVLKQNAKQ--QVDQALQNI--LNAKQALNGDERV---ALAKTNGKHD 650
QY 664 ---NSYALADGEHF-----IVDP-----DKLSSTIVDL-----688
Db 651 IDQLNALNNAQODGFKGRIDQSDNLDNQIQIVDEAKALNRAMDQLSQSEITONEGRKSGT 710
QY 689 -----EKILKLIEGLDYSTLRQTTQ---FRDMTTAGKALLSKSNLRQCE 731
Db 711 NYVNADTVQKVYDVTDKAKQALDKSTQONLTAKQVILKNDAVTAAKKALNGEERLNNR 770
QY 732 KQKFLQ-----BAQFFLGRVLDLDAIAKAEKALVTKKATKNGQLL---ERSINK 777
Db 771 KAEALQRLDQLTHLNNAQRLAIQGINNAETLNKASRAI--NRATKLDNAMGAVQOYIDE 828
QY 778 AVLA-----YNN-----SAIKKANVVKLEKELDLTLGLV 806
Db 829 QHLGVISSTNYINADNKLKANYDNAIAANAHELDKVGNAIAKAEQKQNIIDAQNAL 888
QY 807 EGKGPLAQATMVQGVYLLKTPLEPYEYIYGLNVYFDKSGKLIYALDMSDTTIGEGQKDAYG 866
Db 889 NGDQNLNAK-----DKANAFVNSL---NGLNQOQDOLAH 920
QY 867 NPILNVDE-----DNEGYYHALAVATLADYEGLD-----894
Db 921 KAINNADTVSDVTDIVNNQIDLDNAMEDTLKHLVDNEIPNA---EQVTNYQNADNNAKTNF 977
QY 895 -----IKTIILSKLSQTSIR---QVPTAAYHRAGIFQAIQNAABAEQQLPKPGTHSE 945
Db 978 DDAKRLANTLLNSDNTNYNDINGAIOAVDAIHNLNGDQRLQDAKDKATQISNOALANKL 1037
QY 946 KSSSESANSGKRGLOSN 963
Db 1038 KEIEASNATDQDKLIAKN 1055

RESULT 24
US-09-815-242-5834
; Sequence 5834, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
```

```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5834
; LENGTH: 2437
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-5834

Query Match      4.1%; Score 206.5; DB 9; Length 2437;
Best Local Similarity 19.0%; Pred. No. 0.0026;
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;

QY 46 DYASSGSKKLKINTSGVDDTVTDLFSDKRTTPEKIKONLAKGPRQELKAVNTES 105
DB 543 DHAKQTVSQAHLNNAQKHMEDTLI---DSETRTAVKQDLTEAQAALDQALDQQA 598
QY 106 EKQITSGSQLEQSKESLSLNTKTPSTSNWEICDFITKGNLTVLGSKSGVEK---LSQTDH 162
DB 599 DKDATRASAVYNABPNKQSYDEAVQNAE-----SITAGLNNPTINKGNVSSATQA 650
QY 163 LVLPSQAADGTQLI---QVASF-----FTPDKKTAAEYTSRAGENGESIQDLDVGK 212
DB 651 VISSKNALDGVVERLAQDKQTAGNSLNHLDTLTPAQQAQALNQINNATTRGEVAQKLTQAQ 710
QY 213 EIINGEVFNISVLLKKVTIPTGYKHIGQD-----AFVDNKN---IAEVNLP----- 255
DB 711 ALNQAMEALNSIQDQOQTEAGSKFINKEDPKQDAYQAQVQNAKDLINQTNPTLDKAQV 770
QY 256 BSLETISDYAPAHKALQIDLDNLKATGELAFPNQITGKLSLPRQLMRLAERAFKSNH 315
DB 771 EQLTAQVAKONLHGDKLADKQKHAQVTDL-----NOLGNLNNPQOAL-----BSQ 818
QY 316 IKTIIFRGNLSKVIQASQDNDLSQL--MLPDGLEKIESEAF--NPGDDHYNRNVVL 371
DB 819 INNAATRGVAKLAERAKALQAMQALRNSIQDQOQTESGSKFINKEDPKQDAYQAQV-- 876
QY 372 WTKSKNPSGLATENTYNNPKLSNQESPEIDYTKWLBEDFTYQKNSVTGFSNKGLOKVK 431
DB 877 -----QNAKDLINQTNPTLDKRSQVEQ-----LTQAVTTAKDNL-----HGDQKLA 917
QY 432 RNK-----NLEIPKQ-----HNGVTITEIGNAFRNVDPQNKTLRKYDLEEV 473
DB 918 RDQQAQVTTVNALPNLHAQQAALTDAINAAPRTVEAQHV-----QTATELDHAETL 971
QY 474 KLPSTIRKIGAFAPQSNLKSPEASDDLEEIKEGAF-----DPTNGSNANKDAVDQVL 1026
DB 972 K--NKVDQVNTDKAPNYT---EASTDKKEAVDQALQAESITDPTNGSNANKDAVDQVL 1026
QY 510 -----MNNRIETLEKOKLVITGDAFHINHIYAILVPSVQEIERSAFRONGANN 560
DB 1027 TKLQEKENELNGERNVAEAKTOAKQOTIDQLTLN-----ADQIATA--KQN----- 1070
QY 561 LIFMGSKVKTGEMAFLSNRL-----EHLDLSEKQKLTETPQAFSD--NALK 606
DB 1071 -IDQATKQPTIAELVDQATLQNSMDQLQQAVERNANVEQTVDYT-----QADSDKQAYK 1125
QY 607 EYLLPASLKTIREAPKKNHLKQLEVASALSHAFNALDDNDGDQFQFNKVVVKTTH--H- 663
DB 1126 QAIADA-----ENVLKQNAKQ-QVDQALQNI--LNKAQALNGDSRV---ALAKTNGKHD 1174
QY 664 -----NSYALADGHP-----IVDP-----DKLSSTIVDL----- 688
DB 1175 IDQLNALANAQDGFGRIDQSNLDNLNQIQIIVDEAKALNRAMDQISQEIITNEGRGTGST 1234
QY 699 -----EKILKLIBGLDYSLRQTTQTOQ---FRDMTTAGKALLSKSLNRQGE 731
```

```
DB 1235 NYVNADTVQKVQVDETVDKAKQALDKSTGQNLTKAQVILKNDVATAKKAALNGERLNNR 1294
QY 732 KQKFLQ-----BAQFFLGRVDDDKAIAKAERKALVTTKATQNGQLL---ERSINK 777
DB 1295 KAEALQRLDQLTHLNNARQLAIQIINNAETLNKASRAI--NRATKLDNAMGAVQOYIDE 1352
QY 778 AVLA-----YNN-----SAIKKANVRKLEKELDLTLTGLV 806
DB 1353 QHLGVISSTNYINADNKLKANYDNANAAHELDKVGQNAIAKAEQKQNIIDAQNAL 1412
QY 807 EGKGPLAATWVGYYLLKTPPLPEYYIGLVVFDKSGKLIYALDMSDTTIGEGQKDAYG 866
DB 1413 NGDONLANAK-----DKANAFVNSU---NGLNQOQOOLAH 1444
QY 867 NPILNVDE-----DNEGYPHALAVATLADYEGLD----- 894
DB 1445 KAINNADTVSDTVINNIDNDAMETLKHLDVNEIPNA---EQTVNYQNADDNAKTNF 1501
QY 895 -----IKTILNSKLSQTSIR---QVPTAAHYRAGIFQAIQNAABEQLLPKPGTHSE 945
DB 1502 DDAKRLANTLLNSDNTVNDINGAIOAVNDAIHNLNGDRLQDAKDKAIOQINOALANKL 1561
QY 946 KSSSESSEANSKDRGLQSN 963
DB 1562 KEIEASNATDQDKLIAKN 1579
```

RESULT 25

```
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996
```

```
Query Match      4.1%; Score 206.5; DB 9; Length 6281;
Best Local Similarity 19.0%; Pred. No. 0.0097;
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;

QY 46 DYASSGSKKLKINETSGPVDVTDVTLFSDKRTTPEKIKONLAKGPRQELKAVNTES 105
```

Db 4385 DHAKQTVSOLAHNNNAQKHMEDTLI-----DSETRTAVKQDLTEAQALDQALMDALQOOSTA 4440
Qy 106 EKQITSGSQLESKESLSLNKTVPTSTSNWEICDFITKGNLTVLGSKSGVEK---LSOTDH 162
Db 4441 DKDATRASAYVNAEPNKKQSYDEAVQNAE-----SIIAGLNPTKNGVSSATQA 4492
Qy 163 LVLPQQAADGTOLI-----QVASFA-----FTPDKKTAAIAYTSRAGENGESQLDVGDK 212
Db 4493 VISSKNALDGVRLAQDKQTAGNSLNHLDLTPAQQAQALENQINNATTRGEVAQKLTEAQ 4552
Qy 213 EIINEGEVNSVLLKKVTIPTGYKHIGQD-----AFVDNKN--IAEVNLP-----255
Db 4553 ALNQAMEALRNSIQDQQQTEAGSKFINEDPKQDAYQAAVQNAKDLINQTNPTLDKAQV 4612
Qy 256 ESLETTISDYAFALHAKLQIDLDPNLKAIGELAFDQITGKLSLPQLMRLAERAPKSNH 315
Db 4613 EQLTQAVNAQKDLHGDKLADDDKQHAVTDL-----NQLNGLNPPORQAL-----ESQ 4660
Qy 316 IKTIIEFRGNSLVKIGESFQDNDLSQL--MLPDGLEKIESEAFGT--NFGDDHNNRVVL 371
Db 4661 INNAATRGEVAQKLAERAKALDQAMQALRNSIQDQQQTESSGSKFINEDPKQDAYQAAV--4718
Qy 372 WTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEBEDFTYQKNSVTGFSNKGLOKVK 431
Db 4719 -----QNAKDLINQTNPTLDKRSQVEQ-----LQAVTTAKDNL-----HGDQKLA 4759
Qy 432 RNK-----NLEIPKQ-----HNGVTITEIGDPAFRNVDVFNKTLRKDYDLREV 473
Db 4760 RQOQAVTTVNALPNLNHAQQOALTDAINAAPTRTEVAQHV-----QTATELDHAMEYTL 4813
Qy 474 KLPSTIRKIGAFAPFQSNLKSFEASDDLEEIKEGAF-----509
Db 4814 K--NKVDQVNTDKAQPNYT---EASTDKKEAVDQALQAAESITDPTNGSNANKDAVDQVL 4868
Qy 510 -----MNRITELBLKDLVTIGDAAFHINHIYAIIVLPESVOEIGRSFRONGANN 560
Db 4869 TKLOEKENELNGERNVAEAKTQAKOTIDQTLN-----ADQIATA--KQN-----4912
Qy 561 LIFMGSKVKTGEMAFPLSNRL-----EHLDLSEKQKLTETIPVQAFSD--NALK 606
Db 4913 -IDQATKLPQIAELVDQATQLNQSMDQLOQAVNEHANVEQTVDYT-----QADSQKQAYK 4967
Qy 607 EYLLPASLKTIREEAFKQKHLQLEYASALSHIAFNALDNDGDEQFNDKVVVKT--H- 663
Db 4968 QAIADA-----ENVLKQNAKQ-QVDQALQNI-LNAQALNGDERV---ALAKTNGKD 5016
Qy 664 ----NSYALADGEHF-----IVDP-----DKLSSTIVDL-----688
Db 5017 IDQLNALNNAQQDGFGRIDQSNDLNQIQIIVDEAKALNRAMDQLSQEITDNEGRTKGST 5076
Qy 689 -----EKILKLEGLDYSTLRQTQTO---PRDWTAGKALLSKSNLRQGE 731
Db 5077 NYVNAQTQVKQVYETVDVRAKQALDKSTQGNITAKQVIKLNDVAVTAAKKALNGEERLNR 5136
Qy 732 KQKFLQ-----EAQFPLGRVLDLKAIAKAEKALVTKATKNGQLL---ERSINK 777
Db 5137 KAAELQRLDQTLHNAQQLALQIINNAETLNKASRAI--NRATKLDNAMAQVOYIDE 5194
Qy 778 AVLA-----YNN-----SAIKKANVYKLEKELDLTLGLV 806
Db 5195 OHLGVISSYTNINADNLKANYDNIANAHAHELDKVGQNAIAKAEAEQLKQNIIDAQNAL 5254
Qy 807 ECKGPLAQATWQGVYLLKTPPLPEYYIGLVNYPDKSGKLIYALDMSDTIGEGQKDAYG 866
Db 5255 NGDQNLANAK-----DKANAFVNSL---NGLNQOQOQDLAH 5286
Qy 867 NPILANDE-----DNEGHYHALAVATLADYEGLD-----894
Db 5287 KAINNADTVSDVTDIVNNQIDLDNAMEDTLKHLVDNEIPNA---EQVTYVQNAADDNAKTNF 5343
Qy 895 -----IKTILSKLSQLTISR---QVPTAAVYHRAGIFQAIQNAAAEAFQLLPKPPTHSE 945
Db 5344 DDAKRLANTLLNSDNTNVNDINGAIVQAVNDIAHNLNGDRLQDAKQAIQSIQNALANKL 5403

Qy 946 KSSSESANSKDRGLQSN 963
Db 5404 KEIEASNATDQDKLIAKN 5421
RESULT 26
US-10-470-048B-440
; Sequence 440, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; FILE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
; FILE REFERENCE: SONN:035US
; CURRENT APPLICATION NUMBER: US/10/470, 048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 440
; LENGTH: 10498
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (9728)..(9731)
; OTHER INFORMATION: X = anything
US-10-470-048B-440

Query Match 4.1%; Score 206.5; DB 17; Length 10498;
Best Local Similarity 19.0%; Pred. No. 0.02;
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;
Qy 46 DYAESGSGSKLKNETSGPVDVTVDLPSDKKTTTEKIKONLAKGPREGELKAVNTES 105
Db 8525 DHAKQTVSOLAHNNNAQKHMEDTLI-----DSETRTAVKQDLTEAQALDQALMDALQOOSTA 8580
Qy 106 EKQITSGSQLESKESLSLNKTVPTSTSNWEICDFITKGNLTVLGSKSGVEK---LSOTDH 162
Db 8581 DKDATRASAYVNAEPNKKQSYDEAVQNAE-----SIIAGLNPTKNGVSSATQA 8632
Qy 163 LVLPQQAADGTOLI-----QVASFA-----FTPDKKTAAIAYTSRAGENGESQLDVGDK 212
Db 8633 VISSKNALDGVRLAQDKQTAGNSLNHLDLTPAQQAQALENQINNATTRGEVAQKLTEAQ 8692
Qy 213 EIINEGEVNSVLLKKVTIPTGYKHIGQD-----AFVDNKN--IAEVNLP-----255
Db 8693 ALNQAMEALRNSIQDQQQTEAGSKFINEDPKQDAYQAAVQNAKDLINQTNPTLDKAQV 8752
Qy 256 ESLETTISDYAFALHAKLQIDLDPNLKAIGELAFDQITGKLSLPQLMRLAERAPKSNH 315
Db 8753 EQLTQAVNAQKDLHGDKLADDDKQHAVTDL-----NQLNGLNPPORQAL-----ESQ 8800
Qy 316 IKTIIEFRGNSLVKIGESFQDNDLSQL--MLPDGLEKIESEAFGT--NFGDDHNNRVVL 371
Db 8801 INNAATRGEVAQKLAERAKALDQAMQALRNSIQDQQQTESSGSKFINEDPKQDAYQAAV--8858
Qy 372 WTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEBEDFTYQKNSVTGFSNKGLOKVK 431
Db 8859 -----QNAKDLINQTNPTLDKRSQVEQ-----LQAVTTAKDNL-----HGDQKLA 8899
Qy 432 RNK-----NLEIPKQ-----HNGVTITEIGDPAFRNVDVFNKTLRKDYDLREV 473
Db 8900 RQOQAVTTVNALPNLNHAQQOALTDAINAAPTRTEVAQHV-----QTATELDHAMEYTL 8953
Qy 474 KLPSTIRKIGAFAPFQSNLKSFEASDDLEEIKEGAF-----509
Db 8954 K--NKVDQVNTDKAQPNYT---EASTDKKEAVDQALQAAESITDPTNGSNANKDAVDQVL 9008
Qy 510 -----MNRITELBLKDLVTIGDAAFHINHIYAIIVLPESVOEIGRSFRONGANN 560
Db 9009 TKLOEKENELNGERNVAEAKTQAKOTIDQTLN-----ADQIATA--KQN-----9052

```
Qy 561 LIFMGSKVKTGEMAFSLNRL-----EHLDLSEKQLTEIPVQAFSD--NALK 606
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9053 -IDQATKIQPIAELVDQTLNQSMDQIQAVNEHANVEQTVDYT-----QADSDKQNAIK 9107
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 607 EYLLPASLKTREEAFKKNHLKQLEVASALSHIAFNALDDNDGDQDPKNVVKTH--H- 663
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9108 QAIADA-----ENVLKQNAKQ-QVDQALQNI-LNAQALNGDERV---ALAKTNGKHD 9156
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 664 ----NSYALAGEHF-----IVDP-----DKLSSTIVDL----- 688
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9157 IDQLNALNNAQQGFKGRIDOSNDLNLQIQIVDEAKALNRAMDQLSQBITDNEGRKTGST 9216
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 689 -----EKILKLEGLDYSTLRQTQTQ-----FRDMTTAGKALLSKSNLRQGE 731
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9217 NYVNADTVQVYDETVDKAKQALDKSTGQNIJAKQVTKLNDAVTAAKKALNGBERLNR 9276
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 732 KOKFLQ-----BAQFPLGRVDLDKAIKAEKALVTKKATYKQGLL---ERSINK 777
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9277 KAEALQRLDQLTHLNAQRQLAIQIINNAETLNKASRAI--NRATKLDNANGAVQOYIDE 9334
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 778 AVLA-----YNN-----SAIKKANVKKLEKELDLLLTGLV 806
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9335 QHLGVISSTYINADNKLKANYDNAIAANAHELDKVGQNAIAKAEQKQNIIDAQNAL 9394
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 807 EKGKPLAQTWQGVYLLKTPLPPEYYIGLVNVPDKSKGLIYALDMSDTIGEGQKDAYG 866
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9395 NQDQNLANAK-----DKANAFVNSL---NGLNQOQODLAH 9426
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 867 NPILNVDE-----DNEGYPHALAVATLADYEGLD----- 894
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9427 KAINNADTVSDVDIVNNQIDLNDAMETLKLHLDVNEIPNA---EQTVNYQADNNAKTNF 9483
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 895 -----IKTILNKSLSQTSIR---QVPTAAYHRAGIFQATQNAABAEQLLPKPGTHSE 945
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9484 DKAELANTLLSDNTYNDINGATQAVNDATHNLNGDRLQDQAKDKAIQSIQNALANKL 9543
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 946 KSSSESANSKDRGILQSN 963
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 9544 KEIASNATDQDKLIAXN 9561
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 27

```
US-10-282-122A-70580
; Sequence 70580, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haefelbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
```

```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70580
; LENGTH: 6641
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6636) .. (6636)
; OTHER INFORMATION: X-any amino acid
; US-10-282-122A-70580

Query Match 4.0%; Score 205; DB 15; Length 6641;
Best Local Similarity 19.1%; Pred. No. 0.013;
Matches 201; Conservative 178; Mismatches 447; Indels 224; Gaps 45;

Qy 47 YAESGSKLKLKINTSGV--DDTVTLDFSKRTTPKIKDNLAKGPREBELKAVTENTE 104
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4050 YNOAVNKAANIINDQPTPVMADEIQSVLNEVKQT---KDNL-----HGQKLANDKTD 4100
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 105 SEKOITSGSQLESKESLSLNKTVSTSNMEICDFITKGNLTVLGSKGVEKLSQTDHLV 164
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4101 AQTALNALYLNQAQRGNLETKVQNSNSRSPVKVQLANQLNDAMKLDLALGNDAIK 4160
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 165 LPSQAADQTLQIVASFAFTPKKTAETAYTSRAGEGEISQL-----DVGKEII 215
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4161 QTSNVINEDTSQQVNFDEYTDGKNIVABQTNPNWSPNTINTIADKITEAKNDLHGVQKL 4220
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 216 NEGEVFNLSLLKVT-IPTGYKHIGQDAFVDNKNIAEVN-----LPESLETISDYAF 266
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4221 KQAQQQSINTINQMTGLNQAQKEQLNOEIQOTQTRSEHVQVINKAQAALNDSMTLRQSIT 4280
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 267 AHLAKQIDLPDLNKAIGELAPPN-----QITGKLSLPQLMLAEAFKSNHIKTIE 320
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4281 DEHEVKQTSNYIN-ETVGNQFAYNNAVDVRVKQIINQTSNPTMNPLEVERA-----TSNVK 4334
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 321 PRGNLSKVGESAPQDNDLSQLMLP---DGLEKIESEAFGTG-----NPGDDHNNRV 369
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4335 ISKDALH--GERELNDKNKSKTFAVNHLDNLNQAQKEALTHEIEQATIVSQVNNIYNKAK 4392
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 370 VLWTKSGK-----NPSGLATENTYVNPDP---KSLMQE-----SPEIDYTKM 407
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4393 ALNNDMKKLDIVAQQDNVRSNNYINEDSTPQNNYNDNTINHAQSIIDQVANPTMSHDE- 4451
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 408 LEEDPTYQKNSVTGFSNK-GLQVKRNKMLEIPKHNGVTTITEIGDNFRNVDPQNKTLR 466
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4452 IENAINNIKHAINALDGEHLQQAQKNAUL-----LINSNLNDLNAPORDALNRLVN 4502
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 467 KYDLEEVKLPSTIRKIGAFQSNLKSFEASDDLEEIKEGAFMNNRIETTELKDKLVTI 526
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4503 EAQTR-----KV-----AEQLQSAQALNDAMKHLRNSIQNQ--SSVRQESKYNA 4546
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 527 GDA-AFHINHIYAI-----VLPESVQIEGRSAFRQ-----NGANNLIPMGSKVKTGEMA 575
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4547 SDAKKEQYNH--AVREVENIINEQHTLDEIKQLTDGVNQANNDL-----4591
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 576 FLSNRLHLDLSEKQLTEIP-----VQAFSDNALKE-----VLLPAS 613
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4592 ---NGVELLDADKQNAHQSIPTLMHLNQAQOQNALNEKINNNAVTRTEVAATIGQAKLLDHA 4648
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 614 LKTIRESAPKKNHLKQLEVASALSHPAL--DNDGDGEQFDN-----KVVVKTHNSY 666
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 4649 MENLEESIKKEQVKQ-----SSNYINEDSDVQETVDNAVDHVTILNQTVNPTL 4698
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```



```
/ LENGTH: 10203
/ TYPE: PRT
/ ORGANISM: S.epidermidis
US-10-724-972A-4098

Query Match      4.0%; Score 205; DB 18; Length 10203;
Best Local Similarity 19.1%; Pred. No. 0.024;
Matches 201; Conservative 178; Mismatches 447; Indels 224; Gaps 45;

Qy 47 YAESGSKLKNINETSQGV--DDTVDFDSDKRTTPKIKONLAKGPREQELKAVTENTE 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7618 YQAVNKAHNIINDOPTFMANDEITQSVLNEVKQT---KONL-----HGDKLANDKTD 7668
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 105 SEKQITSSQLEOSKESLSLNTKTPSTSNWEICDFITKGTNTVLGLSKSGVEKLSQTDHLV 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7659 AQATNALNLYNQAGRNLETKVQNSRPEQKVQVQLANQANDAMKLLDALLTNDIAIK 7728

Qy 165 LPSQAADGTQTIQVAFAPDPKTAIAEYTSRAGENGESISOL-----DVGKEII 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7729 QTSWYNEDTSQVNFDEYTDGKNIVAEQTNPNMSPNTNINTIADKI TEAKNDLHGQKL 7788
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 216 NEGEVFNVLKKVT-IFTGYKHIGQDAFVQDNKNIAEVN-----LPESLETISDYAF 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7789 KQAQOQSINTINQMTGLNAQKEQLNQBIQTOTRSEVHQVINKAQLNDSMTLRQSIT 7848
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 267 AHLALKQIDLPNKAIGELAPFDN-----QITGKLSLPQLMLAERAPKSNHIKTI 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7849 DBHEVKQTSNYIN-ETVGNQTAYNNAVDRVKQIINQTSNPTMNPLEVERA-----TSNVK 7902

Qy 321 FRGSLKVI GEASFQNDLSQMLP---DGLEKIESEAFG-----NPGDDHNNRV 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7903 ISKDALH--GERELNDNKNSTFAVNHLDNLNQAQKEALTHEIEQATIVSQVNINYNKAK 7960
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 370 VLMTKSGK-----NPSGLATENTYVNPD---KSLMQE-----SPEIDYTKW 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7961 ALNNDMKKLKQIVAQDQNVQNSNYINEDSTPNQMYNDTINHAQSIIDQVANPTMSHDE- 8019
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 408 LEEDTYQKNSVTGFSNK-GLQVKRNKLEIPKQHGVTITEIGDAPRVDFQNKTLR 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8020 IENAINNKIHALDGBHKQQAQKQENANL-----LINSNDLNAPQORDAINRLVN 8070
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 467 KYDLEVKLPSTIRKIGAFAPQSNLKSFEASDDLKEEGAFMNNRIETLEKDKLVTI 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8071 EAQTE-----KV-----AEQLQAQALNDAMKHLRNSIQNQ--SSVRQESKYINA 8114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 527 GDA-AFHINHIYAI-----VLPESVOETGRSAFRQ-----NGANNLIPFGSKVKTLGEMA 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8115 SDAKKEQYNH--AVREVENIINEQHPTLDEKIIKQITDGVNQANDL-----8159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 576 FLSNLEHLDLSEQKQLETP-----VQAFSDNALKE-----VLLPAS 613
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8160 ---NGVELLDADQKNAHQSIPTLMHLNQAQQNALNEKINNNAVTRTEVAAITGQAKLLDHA 8216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 614 LKTIREAFKQHLKQLEVASALSASHIAFNAL--DDNDGBOQFDN-----KVVKTHNSY 666
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8217 MENLEESIKDEQVQK-----SSNYINEDSDVQETYNADVDHTEILNQVNTPL 8266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 667 ALADGEHFPDVKLSSITVDLEKILKIEGLD--YSTLRQTTQTPQFRDMTTAGKALLSK 724
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8267 SIEDIEHAINEVNAQKQLRGKQKLYQITDLADKELSLDDLTSQSSSSISNQIYVTAKR 8326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 725 SNLRQG-EKQKFLQBAQFLGRVDLDKAIAKAEKALVTTKATYNGQLLERSINKAVLAYN 783
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8327 TEVAQAIKAEKSLNHAM-----KALNKVYKNAKDVLDSSRFNEDQPEKKAYQQAINHVD 8381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 784 NSAIKKAN-----VKRLEKELDLTGLVEGKGPLAQ-----ATMVQGVYLLKTPPLP 831
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8382 SIIHQTPMPDPTVINIETHELETAQNHLGDQKLAHAQQAQDAANVING--LIHLNVAQR 8439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 832 EYIGLNVYFDKSGKLIIVALDMSDTIGEG-----QKDAYGNPILN-----VDEDNEGYHAL 882
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8440 EVMINTNTNATREKVAKNLNDNAQALDKAMETLQQVWAHKNLNDLSKYLNEDSK-YQQQ 8498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
Db 1405 QDILNAQTNDVTQIKDQAVADTQGITADTTIKDVAK-----DELATKAN----- 1449
Qy 175 LIQVASFAPTPDKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFNYSLLKKVTIPTG 234
Db 1450 -----EQKALIAQTADATTEKEQANQOVDAQ-----LTQG 1480
Qy 235 YKHIGQDAFVNDKNIAEVLNPESELTISYAFALHAKQIDLPDNLKAIGELAFPDNQIT 294
Db 1481 NQNIENASQSDVNTAKDNAIOAIDPIQ-----ASTDVKTNARA-ELLTEMQNKIT 1530
Qy 295 GKLSLPRQLMRLAERAFKSHIKTIEFRGNSLKVIGEASFQDNDLSQLMLPDGLEKISE 354
Db 1531 EIL-----NNETTNEEKGNDIGPV-RAAYE-----EGLNINAA 1564
Qy 355 AFTGNPGDDHNNRVVLWTGKGNPSGLATENTYVNPDKSLWQESPEIDYTKWLEBDFTY 414
Db 1565 TTIGD-----VTTAKTAVQKVQQLHANPVKKPACK-----KELDOAAD 1604
Qy 415 QKNSVTGFGKGLQKVRNKNLEIPKOHNGVITTEIGDNAPRVDPQNKTLRYDLEEVK 474
Db 1605 KKTQIBQTPNASQOEINDAQ-EVDTELN-----QAKTNVD-QSSTNEVVD----- 1648
Qy 475 LPSTIRKIGAFQSNLKSFEASDDLSEIKGAFMNNRIETLE--LKDKLVTIGDAAPH 532
Db 1649 -----NAVKEGKAKINAVKTFSEYKKDALKAKIEDA----- 1678
Qy 533 INHIAVLVPESVOEIGRSAFRONGANNLIFMGSKVKTILGEMAFLSNRLEHL-----DL 586
Db 1679 -----YNAKVNEADN-----SNASTSSEIAEAKQKLAELKQTADQNV 1715
Qy 587 SEQKQTEIPVQAFSD-NALKEVLLPASLK-----TIREBAPKK 624
Db 1716 NQATSKODIEVQIHNLDNDINDYTIPTGKESATTDLYAVADQKKNISADTNATQDEQ 1775
Qy 625 NHLKQLE--VASALSHIAFNALDNDGDE-----QEDNKVVVTHHNSVALA- 669
Db 1776 QAIKQVDQNVQTALENIN-NGVDNGVDVDDALTQGKAADITIQVDATVFKPQANQAIKAE 1834
Qy 670 DGEHFTVDPDKLSS--TIVDLEKILKIEGLDYSLTRQTTOTOTQFRDMTTAGKALLSKSL 727
Db 1835 DYKESIDHSDQLTAEKTEALAMIKQITQAKQGITDATTAEVEKAKAQGLEAFDNIQI 1894
Qy 728 ROGEKQKFLQEAQFFLGR-----VLDKRAJAKAEKALVT-----KKATKNGQLLERSIN 776
Db 1895 DSTEKQKAEIELETALDQIEAGVNVVDADATTEKBAFTNALEDILSKATED--ISDQTN 1952
Qy 777 KAVLAYNNSAIAKANKVRL-----EKELDLTLGLVEGKPLAQATWVG 820
Db 1953 AEIATVKNSALEQLKQQRINPVVKNKNALEAIREVNVNKQIEII-----KNADADASAK 2005
Qy 821 VYLLKTPPLPEYIIGLVVYFDK-SGKL-----IYAL-----DMS 854
Db 2006 --IART-----DLGRYFDRFADKLQKTQNTQTEVASELQNTVPAIEAIVPQNDPDAN 2054
Qy 855 DT-IGEGQKDAYGNPILANVDEDNIEGHYALAFADYEGDLKTIILNSKLSQTSIRQVP 913
Db 2055 DTNNGTDNDNATANSANATPENTGQPNVSEIT--DNGRADASPPTPNNSDAATGETTVT 2112
Qy 914 TAAHFRAGIFQAIQNAABAEQLLP-----KPTHSEKSSSSSES 952
Db 2113 SATDDAKDFQANNSADASTNSPTMNDNVTSKPEVESTNNGTTDKPVTETDNATPAES 2172
Qy 953 A---NSKDRGLQSNPKTNRGRHSAILPRTGS 980
Db 2173 TTNNSTTTATNENAPTG---STATAPTAS 2200
```

RESULT 31

US-10-282-122A-43811

; Sequence 43811, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

```
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

```
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
```

```
; CURRENT APPLICATION NUMBER: US/10/282,122A
```

```
; PRIOR FILING DATE: 2003-02-20
```

```
; PRIOR FILING DATE: 2000-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
```

```
; PRIOR APPLICATION NUMBER: 60/206,848
```

```
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
```

```
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: 60/230,335
```

```
; PRIOR FILING DATE: 2000-09-06
```

```
; PRIOR APPLICATION NUMBER: 60/230,347
```

```
; PRIOR FILING DATE: 2000-09-09
```

```
; PRIOR APPLICATION NUMBER: 60/242,578
```

```
; PRIOR FILING DATE: 2000-10-23
```

```
; PRIOR APPLICATION NUMBER: 60/253,625
```

```
; PRIOR FILING DATE: 2000-11-27
```

```
; PRIOR APPLICATION NUMBER: 60/257,931
```

```
; PRIOR FILING DATE: 2000-12-22
```

```
; PRIOR APPLICATION NUMBER: 60/267,636
```

```
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
```

```
; PRIOR FILING DATE: 2001-02-16
```

```
; Remaining Prior Application data removed - See File Wrapper or PALM.
```

```
; NUMBER OF SEQ ID NOS: 78614
```

```
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 43811
```

```
; LENGTH: 6713
```

```
; TYPE: PR1
```

```
; ORGANISM: Staphylococcus aureus
```

```
; US-10-282-122A-43811
```

```
Query Match 4.0%; Score 203; DB 15; Length 6713;
```

```
Best Local Similarity 19.5%; Pred. No. 0.018; Mismatches 396; Indels 376; Gaps 64;
```

```
Matches 232; Conservative 183;
```

```
Qy 34 KQTOASSISGAD-----YAESGSKSLKINETSGP-VDDTVTDLFSDKRTTPEKIKDN 86
```

```
Db 5050 QQTEAGSKFNEFKPKDAYQAAVQNAKDLINQTNPTLDKRSQVQLTOAVTT---AKDN 5106
```

```
Qy 87 L---AKGPREQLKAVTEN-----TESEKQITSGSQLEQSK 120
```

```
Db 5107 LHGDQKLARDQQAQAVTTVNALPNLHAAQQTLTDAINAAPTREVAHQVQTATELDHAME 5166
```

```
Qy 121 SLSLNTKVPSTNSWEICDFITKGNLTVLGSKSGVEKLQSTQHLVLPQAAADCTOLIOWAS 180
```

```
Db 5167 TLK-NKV-----DOVNTDKAQPNTTEASTDKKBAVD-----QALQAAQ 5203
```

```
Qy 181 FAFTPD-----KKTAAEYTSRAGENGISQIDVDGKEIINEGEVFNYSLLKKVTIPTGY 235
```

```
Db 5204 SITDPTNGSNANKDAVEQALTKLOE--KVNEL--NGNERVAEAKTOAKOTIDOLT----- 5254
```

```
Qy 236 KHIGQDAFVNDKN-----IAEV-----NLPESLETISDYAFALHAKQ-IDL----- 276
```

```
Db 5255 -HLNADQIATAKONIDQATKLOPIAELVDQATQLNQMDSQDQQAQVNEHANVEQTIDYTOA 5313
```

```
Qy 277 -PDNLKAIGE-LAFPDN---QITGKLSLPRQLMRL--AERAF-----KSNHIKTIE 320
```

```
Db 5314 DSDKQKAYKQIADAENVLKQNAKQOVDQALQNTILNAKQALNGDERVALAKTKGHID 5373
```

```
Qy 321 FRGNSLKVIGEASF-----QDNDLSQLM-----LPDGLKIESEAPTGNPGDDHNN 367
```

```
Db 5374 QLNALNNAQQGFGKGRIDQSDNDLNIQIQIVDEAKALNRAMDQL-SQBITGNEGR----- 5426
Qy 368 RVLWTGSKNPSGLATENTYVNPDKSLWQSPSE-IDYTKW-----LE 409
Db 5427 -----TKGSTN-----YVADTQVQVYDEAVDKAKQALDKSSQNLTAEQVIKLN 5472
Qy 410 EDFTYQKNSVTGFS-----NKGLQKVKR 432
Db 5473 DAVTAACKALNGEERLNKRKAELQRLDQLTHLNNARQLATQIINNATLTKASRALNR 5532
Qy 433 NKQLS-----IPKOHNGV-----TITEGDNAFRNV-----DFQNKTLRK 467
Db 5533 ATKLDNAGAVQOYIDEHGLGVISSTVYNADDNLKANYDINAANAHELKVKQVNAIAK 5592
Qy 468 YDLERVK-----LPSTIRKIGAPAFQSNL-----KGFSEASDDLEE 503
Db 5593 ABAEQLKQNIIDAQNALNGDQNLAKADKANAFVNSLNGLNQOQDLAKHAINNADTVSD 5652
Qy 504 IKGAPMNRNRIETLKLVTIGDAAAPHINHIYAVLPESVQSIGRSPRQNGANNLIF 563
Db 5653 VTD--IVANQID-----LNDAMETLKLHVDNEIPNABQTVYVYQYQADNNAKTNFD 5700
Qy 564 MGSKVKTL-----GEMAFLSNRLEHLDSEKQ-LTEIPVQAFSD---NALKE 607
Db 5701 AKRLANTLNSDNTVNDINGAIAQVNDAIHNLNGDRLQDAKDAKAIQINQALANKLKE 5760
Qy 608 VLLPASLKTIRREAPKQNLHKLQEVASALSHIAFNALDNDGDEQFD---NKVVVKTHN 664
Db 5761 I--EASNATDQDKLIAKNAEEL-ANSIINN--NKATSNQAVSQVQTAGNHAIEQVHAN 5815
Qy 665 SYALADGEHFIVDPKLSSTIVDLK-ILKLEIGLDYS-----TLRQTTQTQ 710
Db 5816 EIPKAK-----IDANK-----DVKQVQALIDEIDRNPNTLTKSKQALKDRINQILOQ 5864
Qy 711 FRDMTTAGKALSKNLKRGKQ--KFLQEAQFLL-GRVDLDKAIKAEKALVTK----- 762
Db 5865 HNDINNA-----LTKEIEQAKLAQALQDIKOLVKAEDAKQDVQKQVQALIDEIDQNP 5920
Qy 763 -KATKNGQLLERSINKAVLAVNNSAIKXANVRLEKELDLTLGLVEGKGPLAQATMVQGV 821
Db 5921 NLTDKEKQALKDRINQ-ILQOQHNGINNMTK---EEIE-----QAKAQLAQA----- 5964
Qy 822 YLLKTPLPPEYIYGLNVYFDSKGK-LIYALDMSDTIGEGQKDAVGNPILAVDEB--NEG 878
Db 5965 --LKEIKOLVKAENAKQDVQKQVQALIDEIDQNPNTLTKSKQALKDRINQILOQHNDI 6022
Qy 879 YHAL-----AVATLADYEG-L-DIKTLNLSKLSQTSIROVPTAAVHRAGIFQAIONAA 930
Db 6023 NNMATKEIEQAKAQLA--QALQIDKLVKAKEDAKNAIKAL-----AN 6064
Qy 931 ABAEQLPKPGTHSEKSSSS--ESANSKDRGLQS--NPKT-----NRG 969
Db 6065 AKRDQINSNPDLTPBQAKALKKEIDBAEKRALQVYENAGTIDQLNRG 6111
```

RESULT 32

```
US-09-815-242-5703
; Sequence 5703, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
```

```
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5703
; LENGTH: 2025
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2025)
; OTHER INFORMATION: xaa = Any Amino Acid
US-09-815-242-5703
```

```
Query Match 4.0%; Score 202.5; DB 9; Length 2025;
Best Local Similarity 18.5%; Pred. No. 0.0035;
Matches 216; Conservative 174; Mismatches 410; Indels 367; Gaps 55;
```

```
Qy 16 SVVTHNQEVSLVKEPILKQTOASSISGADYAESS--GKSKL-----KINET 61
Db 413 AAVQHAQDLINQTSNPTLKDQAVBQLTQGVNQAKNLHGDQKLAADDKQHAVTDNLQNSL 472
Qy 62 SGPVDDTVTDLFSKRTTPEKIKDNLAKGPR-EGELKAVTENTSEKQITSGSOLQESKE 120
Db 473 NNPOQALAESINNAATDE-VAQKLAEQAQALDQAMQALRNSIODQOQTESGSKF----- 526
Qy 121 SLSLNTKVPSTSNWEIC-----DFITK-GNTLVGLSKSGVKEKLSQT-----DHL----- 163
Db 527 ---INEDPKQDAYQAAVQHAQDLINQTSNPT--LDKSQVBQLTQAVTTAKDNLHGDQKL 581
Qy 164 -----VLP-----SQADGTQLIQVASFAPDKKTAIAEYTSRAGE----- 200
Db 582 ARDQQAQVTVVNALPNLNLHQAQQALTDANAA-----PTREVAQHVQTATELDHAMET 635
Qy 201 -NGEISOLD-----VDGKEIINEGEVFNLSVLLKKVTIPTGYKHIGQDAFVD-- 245
Db 636 LKNKVDQNTDKAQPNYTEASTDKKEAVDQA-----LQAEISITDPTGNSNANKDAVEQAL 691
Qy 246 -----NKNIAEVLNLPESLETISDYAFAL-----ALKQIDLDPNLKAIQEL 286
Db 692 TKLQKQVNLNGNERNVAREAK-AQAKQITDQ--LAHLNADQIATAKQNIQDQATKLPQIAEL 748
Qy 287 AFPDQNI-----TGKLSLPROLMRLAERAPKSNHIK-- 317
Db 749 VDAQTLQNSMDQLQAVNEHANVEQTVDYTQADSDKQNAKYQKQIAEAVNLKQNSKQ 808
Qy 318 -----TIEFRGNSLKVIGESFQDNDLSQMLPDCGLEKIESEAFPTGNPGDDHYN 366
Db 809 VDQALQNLNKAQALNGDERVALAKTNKG-HDIDQL---NALNNAQDQGFGRIDQSDHL 864
Qy 367 NRVVLWTKSGK-----NPSGLATENTYVNPDKSLWQSPSE-IDYTKW--- 407
Db 865 NQIQIIVDEAKALNRAMDQLSQEISGNEGRKTSNTYVADTQVQVYDEAVDKAKQALD 924
Qy 408 -----LEEDFTYQKNSVTG--FSNKGKQKVKR-----NKMLEIPK 440
Db 925 KSTGQNTAEQVILKNDVATAAKKALNGEERLNKRKSEALQRLDQLTHLNNARQLAQ 984
Qy 441 QHNGVTITEIGD--NAFRNVDFQNKTLRKY--DLEEVKLPSTIRKIGAPAFQSNLKS--- 494
```

```
Db 985 INNAETLNKASRAINRATKLDNAMGAVQOYIDBQHLGVISSTNYINA-----DDNLKANYD 1040
QY 495 ---FEASDDLEIEKEGAFMNNRIETLEKDKLVITIGDAAAFHINHIYAIVLPSVQSIGRS 551
Db 1041 NAIANAHELDKVGNAIA--KAAEQKQNIIDAQNAL-----1077
QY 552 AFRONGANNLIPMGSKVKTIGEMAFISNRLEHLDLSEQKQLTEIPVQAFSDNALKEVLLP 611
Db 1078 ---NGDONLANAKDRA--AFVN-----SLNGLAQ-----1102
QY 612 ASLKTIREAFKQKHLKQLEVASALSIAFNALDNDGDGEQF-----DNKVVVVKTHNSYA 667
Db 1103 ---QOQDLAHNAINNADTVSDVTIDVNNQIDLDNDAETLKLHLDNINPNAEQVTNYQ 1156
QY 668 LADGEHFIVDPDKLSSTIVLEKILKLEGLDYSTLRQTOTQFRDMTTAGKALLSKSNL 727
Db 1157 NAD-----DNAKTNFDDAKELA-----NTLLNSDNTNVNDINGAIQAVKDAION 1200
QY 728 ROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTK-----KATKNGOLLERSINKA-V 779
Db 1201 LANGEQR--LQEAQ-----DKAIQNVNKLADKLKEIEASNATDQDKLIAK--NKABE 1248
QY 780 LAYNSAIIKANVRLKELDLTLGLVEGKPLAQATMVQGVYLLTPLPEYVYIGLVN 839
Db 1249 LA--NSIIN--NINKATSNQDSVQVQTAGNQAIEQVHANE-----IPKAKIDANK 1294
QY 840 YFDKSGK-LIYALDMSDTTIGEGQDAYGNPILNVDED--NEGHALAVATLADYEG--- 892
Db 1295 DVDKQVOALIDEIGRPNLTDEKQKALDRINGILOQGHNDINNALTKAEIQAERLQ 1354
QY 893 --LDIKTIILNSKLSQTSIRQVPTAAHYRAGIFQAIQNAEAEQLLPKPGTHSEKSSS 950
Db 1355 ALQDIKDLVKAKEDAKNKIKAL-----ANAKRQDINSNPDLTPEQKAKA 1398
QY 951 --ESANSKDRGLQS--NPKT-----NRG 969
Db 1399 LKEIDEAEKRALQNVENATQIDQLNRG 1425
```

RESULT 33

```
US-09-815-242-12611
; Sequence 12611, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 12611
; LENGTH: 3158
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-12611

Query Match 4.0%; Score 202.5; DB 9; Length 3158;
Best Local Similarity 18.5%; Pred. No. 0.0065;
Matches 216; Conservative 174; Mismatches 410; Indels 367; Gaps 55;

QY 16 SVVTHNQEVPSLVKPEILKQTOASSISGADYAES--CKSKL-----KINET 61
Db 1544 AAVQAKOLINTSPTLDKQAEQVLTQGVNAOKNLHGDQKLADQKHAVDTDLNQLNSL 1603
QY 62 SGPVDDFTVDLPDSKRTTPEKIDNLAKGPR--EQLKAVTENTSEKQITSGSOLQSKE 120
Db 1604 NNPQQALESQINNAATRDE--VAQKLAEAQALDQAMQALRNSIQDQOQTESSESKF 1657
QY 121 SLSLNTVPSTSNWISIC-----DFITK--GNTVLVGLSKGVEKLSQT-----DHL 163
Db 1658 ---INEDKPKDAYQAAVQAHKDLINQTNPT--LDKSOVEQLTQAVTTAKDNLHGDQKL 1712
QY 164 -----VLP-----SQAADGTQLIQVASPAFTPKKTAIAEYTSRAGE----- 200
Db 1713 ARDQQAQVTTVNALPNLHAQQAALTDAINAA-----PTRTEVAQHVTATDELHAMET 1766
QY 201 -NGEISOLD-----VDGKEINEGEVFNFSYLLKKVTIPTGYKHIQODAFVD-- 245
Db 1767 LKNKVDQVNTDKAQPNYTEASTDKKEAVDQA-----LQAAESITDPTNGSNANKDAVEQAL 1822
QY 246 -----NKNIAEYNLPESLETISDYAFAHL-----ALKQIDLPNLKAIGEL 286
Db 1823 TKLQEKVNLNGNERNVABAK--AAKQITDQ--LAHLNADQIATAKQNIQATKIQPIAEL 1879
QY 287 AFPDNQI-----TGKLSLPLQMLRAERAFKSNHIK-- 317
Db 1880 VDAQTLNQSDMDLOQAQVNEHANVEQTVDTQADSDKQNAKQAEAEVNLKQNSKQ 1939
QY 318 -----TIEFRGNSLKVIGASFOQNDLSQMLPDGLEKTESAFTNGPDHYN 366
Db 1940 VDAQALQNLINAKQALNGDERVALAKTNGK--HDIQDQ--NALLNAAQDQFGKRIDQSHDL 1995
QY 367 NRVLWTKSGK-----NPSGLATENTYVNPDKSLMQESPE--IDYTKW--- 407
Db 1996 NQIQIIVDEAKALNRAMDOLSOEISGNEGRTKSTNYVNAQVQVYDEAVDKAKALD 2055
QY 408 -----LEEDFTYQKNSVTG--FSNKGLOKQVR-----NKNLEIPK 440
Db 2056 KSTGQNLTAEQVILKNDVAATAAKALNGEERLNKRKSEALQRLDQLTHLNNARQLAIQ 2115
QY 441 QHNGVTITEIGD--NAFRNVDFQNKTLRKY--DLEEVKLPSTIRTKTGAFAPQSNNLSK 494
Db 2116 INNAETLNKASRAINRATKLDNAMGAVQOYIDBQHLGVISSTNYINA-----DDNLKANYD 2171
QY 495 ---FEASDDLEIEKEGAFMNNRIETLEKDKLVITIGDAAAFHINHIYAIVLPSVQSIGRS 551
Db 2172 NAIANAHELDKVGNAIA--KAAEQKQNIIDAQNAL-----2208
QY 552 AFRONGANNLIPMGSKVKTIGEMAFISNRLEHLDLSEQKQLTEIPVQAFSDNALKEVLLP 611
Db 2209 ---NGDONLANAKDRA--AFVN-----SLNGLAQ-----2233
QY 612 ASLKTIREAFKQKHLKQLEVASALSIAFNALDNDGDGEQF-----DNKVVVVKTHNSYA 667
Db 2234 ---QOQDLAHNAINNADTVSDVTIDVNNQIDLDNDAETLKLHLDNINPNAEQVTNYQ 2287
QY 668 LADGEHFIVDPDKLSSTIVLEKILKLEGLDYSTLRQTOTQFRDMTTAGKALLSKSNL 727
Db 2288 NAD-----DNAKTNFDDAKELA-----NTLLNSDNTNVNDINGAIQAVKDAION 2331
QY 728 ROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTK-----KATKNGOLLERSINKA-V 779
Db 2332 LANGEQR--LQEAQ-----DKAIQNVNKLADKLKEIEASNATDQDKLIAK--NKABE 2379
```



```

; APPLICANT: Ni et al.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481D1
; CURRENT APPLICATION NUMBER: US/10/994,726
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: 09/830,230
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: Patent in Ver. 3.3
; SEQ ID NO 321
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (573)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (627)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (735)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-994-726-321

Query Match          3.9%; Score 200; DB 18; Length 1119;
Best Local Similarity 19.6%; Pred. No. 0.0022;
Matches 212; Conservative 164; Mismatches 358; Indels 347; Gaps 54;

QY 18 VTHNGEVFLVKEPILKQTOASSISGADYAESGSKLKNINETSQPVDDTVTLFSDKR 77
DB 188 VNDQKNLFNLEK---LKK-----NLGKSNSEILNDSQKIB-----NDKQ 225

QY 78 TT---PEKIDKLNKXGPREQ-----LKAVENTESEKQITSGSOLEQSK-----BS 121
DB 226 NTNLSEKENSEILKTPDNSKYSNNNTTSLKKISSNQKESLSPSPSTIIGKIYRPYS 285

QY 122 LSLINKTPVSTSNWEICDFTTKGNLTG-----LSKSGV-EKLSOTDHLVLPQAADGTO 174
DB 286 YLIKKEK-----YEILDDINTGRTVTLGKNRLKELIKKLSNFKQKVELIENSKKEASN 340

QY 175 LI-----QVASFAPTPDKKTAIBYTSRAGENGESIQLDVDGKEIINEGEVFNYSY 224
DB 341 LLLTLIKKIDIEPNLINIPKPYKK-----EIFQDKEDKK-----PQY 378

QY 225 L-----LKKVIPTGYKHIGQAFVD-----NKNIAEVLNPSLEPI 261
DB 379 LEDLSKVHSIKPIDLENTKSR--QQAIDKLNFLKNPNPDAQASKTLAQANKIQHLEDL 436

QY 262 SDYAFHAHLKOIDLPDNLK---AIGELAFF-----DNOITGKLSLPRQLMRLAERAF 311
DB 437 KSKVH---SIKPIDL-ENTKSRQAQKIDLNFLKNPNPDAQASKTLAQANKIQHLEDLS 492

QY 312 KSNHIKTIFF-----RGNLSKVIGEASFQND-----LSQLMLPDGKEISEAPT 357
DB 493 KVHSIKPIDLENTKSRQAQKIDLNFLKNPNPDAQASKTLAQANKIQHLEDLSKVHS 550

QY 358 GNPQDDHNNRVVLWTKSGKNSGLATENTYNNPDKSLWQESPEIDYTKWLEEDFTYQKN 417
DB 551 IKPID-----LENT-----KSRQAQKIDLN-----EPX-KNN 576

QY 418 SVTGFSGKLGQKVKRNQNLKLEIPKQHNQVTTITEIGNAFNRVDFQNKTLKRYOLEEYKPLS 477
DB 577 PNDQAASKTLAQANKIQHLEDLSK-----VHSIKPIDLENTKSRQAQKIDLN----- 624

QY 478 TIRKIGAFAPQSNLKSFEASDDLEIEKEGAFMNRRIETLE-LKOKLVITIGDAAFHINH 536
DB 625 -----EFXKNPNPDAQASKTLAQ-----NKIQHLEDLSK-----VHSI 659

QY 537 YAVLV-----PESVOEIGRSFRONGANNLIFMGSKVKTLCGEMAFSLNRLHDLSEOK 590
DB 660 KPIDLENTKSRQAQKIDL--NEFLKNPNP-----AQASKTLAQ-----ANKIQHLEDLSK 709

QY 591 QLTEIPVQAFSDNALKEVLLPASLKTIRREAFKQNHLSKQLEVASALSHTAFNALDDNDGD 650
DB 710 VHSIKPIDLENTKSRQ-----AIKDLNE--FXKNPNPDAQASKTLAQ--AY-----ENNED 757

QY 651 -----EOPDNKVVVKT-----HHN-----SYALADGHEFI-----VDPDKLSS----- 683
DB 758 LKKAENAYEKIILKTQEDHYKLGIIIRFKLKKYEHSESFDQTIKLDPKHKKALHNKGI 817

QY 684 TIVDLEKILKLEGLDYSLRTQTQTPRDMTTAGKALLSKSLNLRQGEKQKFLQEAQ--- 740
DB 818 ALMMLNKKKKAIESFE-----KAIQIDKNYGTAYYQKGIAEEKNGD 858

QY 741 -----PFLGRVLDLKAIAKAFKALVTTKATKNGQLLERSIN-----KAVLAYNNS 785
DB 859 MQQAFASFKNAVLDKNPNYALKAGIVSNLGNFKQSEYLNFFNANAKKPEIAIYNLS 918

QY 786 AIKKNVRLKELDLTLGLVE-----GKQPLAQATWVGQVYLLKTP 828
DB 919 IAKFEN-NKLESLETKAIDLNPKESEYLYLKASINLKKENYQNAISLYSLVIEKPN- 976

QY 829 PLPEYVIGVNVYFDKSGKLIIVALD-MSDTIGBGQDAYCN----- 867
DB 977 ENTSAYINLAKAYESGKNSQSAISTELEKLNKNKLNALNGLYKKEKNYOKAIEIFEK 1036

QY 868 FILNVEDNEGYHALAVATLADYEGLDIKTIILNSKLSQTSIRQVPTAAYHRAIGFQAIO 927
DB 1037 AIIN--SDIEAKYNLATTILIEINDNTRAKDLR-----EYTKLPNNPEALHALGIIIEYNE 1090

QY 928 N 928
DB 1091 N 1091

RESULT 36
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Betell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394CI-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-932-183A-2

Query Match          3.9%; Score 199.5; DB 9; Length 2285;
Best Local Similarity 20.9%; Pred. No. 0.0063;
Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;
```

```
QY 81 EKIKDNIAKPREQELKAVTENTES-EKQITSGSQLESKESLSLN----- 125
Db 771 KKAQDDPEQS-QQTNVEAITNKDSTDKLIQQYKELQKVKSRSLSDBEQEYLQVTOOL 829
QY 126 -KTVPSSTNWEICDFTTKGNTLVGLSGVKEKLSQT-DHLVLPQAA-----DGTQ 174
Db 830 AQTFPAL----VKGYSQGNAILKTNKELEKAIENKTEYLAKKQETRDSAKKTFEDASK 885
QY 175 LIQVASFAPDPKKTAAIAYTSRAGENGESIQLDVDGKEIINEGEVFNYSLLKKVTIPTG 234
Db 886 EIKKSKDELQYKQ--IADYNDKGRPKWDLIADDDYKVAADKAK--QSMUKAQSDIESG 941
QY 235 YKHIGQDAFVDMKNIAEYNLPESLET-ISDYAFALHALKQIDLPNLKAIAGELAFFDNQI 293
Db 942 NAKVKDSVLSTANAYSSIDISNTLKTISD-----VVKLNKLDLDP-ELEKFSSSL 994
QY 294 TGLKSLPRQLMRLAERAFKSNHIKTIEFRGNSLKVIGBASFQDNDLSQMLPDGLEKIES 353
Db 995 -GKLQEKMQ-----KALDSGDEKAFD--NAKK-----DLQSL----- 1024
QY 354 EAFGNPGDDHYNNRVVLWTSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFT 413
Db 1025 -----ETYSKSDS-----IDVFKMSFD--K 1043
QY 414 YQKNSVTGFSNKGLOKVRKNKNLEIPKOHNGVTITEIGDNAFRNVDFONKTLRKYDLEEV 473
Db 1044 AQKNIKOG--DKLSLSVSKSEVG-----DLGETLAEAGNEA--EDFGK-----LKEA 1086
QY 474 KLPETIRKIGAFAPFOSNNLKSFEASDDLBEETKEGAFMNNRIETLELKDCLVTIGDAAPHI 533
Db 1087 LDANSVDDIKAAIKEMSDAMQF--DSVQDVLNGDIFNN-----TKQVAPLND----- 1132
QY 534 NHIYAVLPESVQIGBSAFQNGANNLI PMGSKVKTGEMAFLSNRLEHLDLSEKQKLT 593
Db 1133 -----LLEKMAE-GKS-ISANEANTLI---QKDKELAQAIISIEGVVVKINRDEVIKQR 1180
QY 594 EIPVQAFSD-----NALK-----EVLPLPASLKTIRRE----- 620
Db 1181 KVKLDAYNDMVTYSNKLKMKTEVNNAIKTLNADTLRIDSLKLRKERKLDMSAEALSLEEV 1240
QY 621 -----AFKQNHKLKOLE-----VASAL-SHIAFNALDDNDGDEQFD 654
Db 1241 KSINNVAADAKKELKLEKMLQPGGYSNSQTEAMQSVKSALESYI--SASEATSTQEMN 1298
QY 655 NKVVVTHNSYALADGEHFIVDPDKLSSTIVD-----LEKILKLIBGL-----DYSTLR 704
```

RESULT 37

```
US-10-927-615-2
; Sequence 2, Application US/10927615
; Publication No. US20050009146A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394CL-US
; CURRENT APPLICATION NUMBER: US/10/927,615
; PRIOR FILING DATE: 2004-08-25
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-927-615-2
```

```
Query Match 3.9%; Score 199.5; DB 17; Length 2285;
Best Local Similarity 20.9%; Pred. No. 0.0063;
Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;
```

```
QY 81 EKIKDNIAKPREQELKAVTENTES-EKQITSGSQLESKESLSLN----- 125
Db 771 KKAQDDPEQS-QQTNVEAITNKDSTDKLIQQYKELQKVKSRSLSDBEQEYLQVTOOL 829
QY 126 -KTVPSSTNWEICDFTTKGNTLVGLSGVKEKLSQT-DHLVLPQAA-----DGTQ 174
Db 830 AQTFPAL----VKGYSQGNAILKTNKELEKAIENKTEYLAKKQETRDSAKKTFEDASK 885
QY 175 LIQVASFAPDPKKTAAIAYTSRAGENGESIQLDVDGKEIINEGEVFNYSLLKKVTIPTG 234
Db 886 EIKKSKDELQYKQ--IADYNDKGRPKWDLIADDDYKVAADKAK--QSMUKAQSDIESG 941
QY 235 YKHIGQDAFVDMKNIAEYNLPESLET-ISDYAFALHALKQIDLPNLKAIAGELAFFDNQI 293
Db 942 NAKVKDSVLSTANAYSSIDISNTLKTISD-----VVKLNKLDLDP-ELEKFSSSL 994
QY 294 TGLKSLPRQLMRLAERAFKSNHIKTIEFRGNSLKVIGBASFQDNDLSQMLPDGLEKIES 353
Db 995 -GKLQEKMQ-----KALDSGDEKAFD--NAKK-----DLQSL----- 1024
QY 354 EAFGNPGDDHYNNRVVLWTSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFT 413
Db 1025 -----ETYSKSDS-----IDVFKMSFD--K 1043
QY 414 YQKNSVTGFSNKGLOKVRKNKNLEIPKOHNGVTITEIGDNAFRNVDFONKTLRKYDLEEV 473
Db 1044 AQKNIKOG--DKLSLSVSKSEVG-----DLGETLAEAGNEA--EDFGK-----LKEA 1086
QY 474 KLPETIRKIGAFAPFOSNNLKSFEASDDLBEETKEGAFMNNRIETLELKDCLVTIGDAAPHI 533
Db 1087 LDANSVDDIKAAIKEMSDAMQF--DSVQDVLNGDIFNN-----TKQVAPLND----- 1132
QY 534 NHIYAVLPESVQIGBSAFQNGANNLI PMGSKVKTGEMAFLSNRLEHLDLSEKQKLT 593
Db 1133 -----LLEKMAE-GKS-ISANEANTLI---QKDKELAQAIISIEGVVVKINRDEVIKQR 1180
QY 594 EIPVQAFSD-----NALK-----EVLPLPASLKTIRRE----- 620
Db 1181 KVKLDAYNDMVTYSNKLKMKTEVNNAIKTLNADTLRIDSLKLRKERKLDMSAEALSLEEV 1240
QY 621 -----AFKQNHKLKOLE-----VASAL-SHIAFNALDDNDGDEQFD 654
Db 1241 KSINNVAADAKKELKLEKMLQPGGYSNSQTEAMQSVKSALESYI--SASEATSTQEMN 1298
QY 655 NKVVVTHNSYALADGEHFIVDPDKLSSTIVD-----LEKILKLIBGL-----DYSTLR 704
```

```
Db 1299 KOALVEAGTSLNWTDOEKANEETKTSMYVVDKYKEALEKVNAAEIDKYNKQVNDYPKYS 1358
QY 705 QTTOTQFRD-MTTAGKALLSKSNLROGEKQKFLQEAQFFLGRVLDLDAKAKAEKALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQOKKLMQ-EQAKLLQD-QIKSGNI-----TOYGIIVTST 1404
QY 764 ATKNG--QLLERSINKAVLAYNNSAIKKNVRLKLEKELDLTLGLVEGK-GPLAQATWVG 820
Db 1405 TSSGGTSPSTGSGYSKYSSYNSAASKYNV-----DPALIAAVIQOESGFNAKARSVG 1459
QY 821 VYLLKTPPLPEYIYGLNVYFDKSKGLIYALDMSDTIGEGQK-DAYGNPILNVDENEGY 879
Db 1460 AMGLQMLPATAKSLGWNAYDP-----YQNVMGGTKYLAQOLEKFGG---NVEKALAA 1511
QY 880 HALAVATLADYEGLD-----IKTIL---NSKLSQLTSLRQVPTAAVHRAGIFOAION 928
Db 1512 NA-GPGNVIKYGGIPPFKETQNYVKIMANYSKSLSSATS-----SIASY 1556
QY 929 AAAEAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG--SFY 986
Db 1557 -----TNNSAFRVSSKYQOESGLRSSP--HKGTDPAAKAGTAIKSLQSGKV 1601
QY 987 GILGYTSTA 995
Db 1602 QIAGYSKTA 1610

RESULT 38
US-10-927-590-2
; Sequence 2, Application US/10927590
; Publication No. US2005059112A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/10/927,590
; PRIOR FILING DATE: 2004-08-25
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PR1
; ORGANISM: Bacillus subtilis
US-10-927-590-2

Query Match 3.9%; Score 199.5; DB 17; Length 2285;
Best Local Similarity 20.9%; Pred. No. 0.0063;
Matches 215; Conservative 155; Mismatch 356; Indels 303; Gaps 54;

QY 81 EKIKNLAKPGREBELKAVNTES-EKOITSGSQESKESLSLN----- 125
Db 771 KKAKDDFEGS-QOTVEAITTNDKTDKLIQYKELQKVKERSITSDEEQLVQTOOL 829
QY 126 -KTPVSTSNWEICDFTTKGNTLVGLSKGVEKLSQT-DHLVLPQAA-----DGTQ 174
Db 830 AQTFPAL----VKGYDSQGNALTKNKELEKAIENTKEYLALKQETRDSAKKTFEDASK 885
QY 175 LIQVASFAFTPKKTAIATYSRABENGHISOLDVGGKEINEGEVFNISYLLKVTIPTG 234
Db 886 EIKSKDELKQYKQ--IADYNDKGRPKWDLIADDDDDYKVAADKAK--QSMKLAQSDIESG 941
QY 235 YKHIGQDAFVDNKNTAEVNLPSLET-ISDYAFALHAKQIDLPNLKAIAGELAFEDNQI 293
Db 942 NAKVKDSVLSTANAYSSDISNTLTSISD-----VNMKMLKDDLOP-EELEKFPSSSL 994
QY 294 TGKLSLRLQMLRAERAFKSNHIKTIETFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIES 353
```

```
Db 995 -GKLEQKMQ-----KALDSGDEKAFD--NAKK-----DLQSL----- 1024
QY 354 EAFITGNPGDDHNNRNVVLWTSGKNPSGLATENTYVNDPKSLWQSPEDIDYTKWLEEDFT 413
Db 1025 -----ETYSKSDSS-----IDVFKMSFD--K 1043
QY 414 YOKNSVTGFSNKGLOKVKRNKNLEIPKQHNGVVTIITEIGDNAPRNVDFOKNTLRKYDLEEV 473
Db 1044 AQKNIKDG--DKSLSSVSEVG-----DLGETLAEGNEA--EDFGK-----LKEA 1086
QY 474 KLPSTIRKIGAPAFOPSNLKSFEASDDLEETKEGAPMNRIBTELKOKLVITIGDAAFHI 533
Db 1087 LDANSVDDIKKAIKEMSAMQF--DSVQDVLNGDIFNN-----TKQVAPLND----- 1132
QY 534 NHIYAIVLPESESQETGRSAPRONGANNLIFMGSKVKTIGEMAFPLNRLHLDLSEKQJLT 593
Db 1133 -----LLEKMAE-GKS- ISANEANTLI--QDKELAQAISIENGVVVKINRDEVIKOR 1180
QY 594 BIPVQAFSD-----NALK-----EVLLPASLKTIRE----- 620
Db 1181 KVKLDAYNDMVTYSNKLAKMTEVNNAIKTLNADTLRIDSLKLRKERKLDMSAEISDLEEV 1240
QY 621 -----APKNNHLKOLE-----VASAL-SHIAFNALDNDGDEQPD 654
Db 1241 KSINNVAADAKKELKLEKMLQPGGYSNSQTEAMQSVKSALESYI--SASEEATSTQEMN 1298
QY 655 NKVVVYKTHNSYALADGEHFIVDPDKLSSTIVD-----LEKILKLEGL-----DYSTLR 704
Db 1299 KOALVEAGTSLNWTDOEKANEETKTSMYVVDKYKEALEKVNAAEIDKYNKQVNDYPKYS 1358
QY 705 QTTOTQFRD-MTTAGKALLSKSNLROGEKQKFLQEAQFFLGRVLDLDAKAKAEKALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQOKKLMQ-EQAKLLQD-QIKSGNI-----TOYGIIVTST 1404
QY 764 ATKNG--QLLERSINKAVLAYNNSAIKKNVRLKLEKELDLTLGLVEGK-GPLAQATWVG 820
Db 1405 TSSGGTSPSTGSGYSKYSSYNSAASKYNV-----DPALIAAVIQOESGFNAKARSVG 1459
QY 821 VYLLKTPPLPEYIYGLNVYFDKSKGLIYALDMSDTIGEGQK-DAYGNPILNVDENEGY 879
Db 1460 AMGLQMLPATAKSLGWNAYDP-----YQNVMGGTKYLAQOLEKFGG---NVEKALAA 1511
QY 880 HALAVATLADYEGLD-----IKTIL---NSKLSQLTSLRQVPTAAVHRAGIFOAION 928
Db 1512 NA-GPGNVIKYGGIPPFKETQNYVKIMANYSKSLSSATS-----SIASY 1556
QY 929 AAAEAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG--SFY 986
Db 1557 -----TNNSAFRVSSKYQOESGLRSSP--HKGTDPAAKAGTAIKSLQSGKV 1601
QY 987 GILGYTSTA 995
Db 1602 QIAGYSKTA 1610

RESULT 39
US-10-926-729-2
; Sequence 2, Application US/10926729
; Publication No. US2005010668A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/10/926,729
; CURRENT FILING DATE: 2004-08-25
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
```


Db 2120 EVIKASPKVGPAPKVCPTPIKEDKQEVKVVKELP-NTGSEMDLPLKELALITGAALL 2178

RESULT 42

US-11-020-509-2

; Sequence 2, Application US/11020509

; Publication No. US20050106648A1

; GENERAL INFORMATION:

; APPLICANT: FOSTER, Timothy et al.

; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .

; FILE REFERENCE: P07263US02/BAS

; CURRENT APPLICATION NUMBER: US/11/020,509

; CURRENT FILING DATE: 2004-12-27

; PRIOR APPLICATION NUMBER: US 10/172,502

; PRIOR FILING DATE: 2002-06-17

; PRIOR APPLICATION NUMBER: US 60/298,098

; PRIOR FILING DATE: 2001-06-15

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 2189

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-11-020-509-2

Query Match 3.9%; Score 199; DB 19; Length 2189;

Best Local Similarity 19.7%; Pred. No. 0.0064;

Matches 201; Conservative 176; Mismatches 359; Indels 284; Gaps 53;

Qy 16 SVVTHNQEVSLVKEPILKQTOQSSSI-SGADYAESGSKLKNINETSPPVDDTVDLFS 74

Db 1232 AINTLNQROQAIES--IKQANTNAEVQAAATVAENN-----IDAVQVDVVK 1276

Qy 75 DKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSKS-----SLSLNTK--- 127

Db 1277 -KQAAARDKITAEVAK--RIEAVKQTPNATDEBKQ-AAVNQINQLKQAFNQINQNTNDQ 1332

Qy 128 VPSISNWEI-----CDPIYKGNVLGLSGVGEKLSQTHLVLPSSQAADGTQLIQVAS 180

Db 1333 VDAATNQAINADNVAEAEVWPKPAIDIEKAVKEKQOQIDNSL----DSTONEK--EVAL 1387

Qy 181 FAFAPDKKTAIAEYTSRAGEISEISQDVGKKEINEGEVFNLSLLKKVT--IPRGYKHI 238

Db 1388 QALAKEKEKALAA-IDQATNSQVNNQAATNGVSAIK-----IIQETKIKPAARSKI 1438

Qy 239 QGDAPVDNKNIAEAVNLPE-----SLETISDYFAFAHLAKQIDLPDNLKAIGELAFF 289

Db 1439 NQKA---NELRAQINQDKAEATAEERQAALDKIND----LVAKAMTNITNDRNTNQVNSD 1490

Qy 290 DNOITGKLSL--PROLMRLAER-----APKSNHIKTIIEFRGNSLKVIGEASFQDND-- 338

Db 1491 TNQALDDIALVTPDHIIVRAAARDVAKQOYEAKEKHEIEQAEHATDBEKQVALNQLANNEK 1550

Qy 339 ----LSQMLPDGLBKIESEFTGNPGDDHY-----NNRVVLWTKSGKNP 379

Db 1551 ALONINQAIANDVGRVSENGIATLKGVPHIVVPEAQEAIAKASADNQV-----ESIKDT 1606

Qy 380 SGLATE-----NTYVNPDKLSQWESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVK--- 431

Db 1607 PHATTDELDEANQIN--DTLKQGGQDIDNT--TQDAAV--NDVRNQTIKAEIQKPKV 1659

Qy 432 -----RNKN-----LEIPKQNGVTIITEIGD--NAFRNVDQNTKLRYKD 469

Db 1660 RRKRAALDNI DESNNQLDAIRNTLDUTTTQDERNVAIAALKNIVNAKNDIAQNKTNAEVD 1719

Qy 470 LEE-----VKL-----PSTIRKIGAFQSNLKSFEASD--DLBEEIKEGA----- 508

Db 1720 QTEADGNNNIKVILPKVQKPAARQSVSAKAEQNAL--IQSDLSSTEEERLAAKHLVEQ 1777

Qy 509 FMNRIETLELKDVLVTIGDAAFHINHIIYAILVP-----ESVQIGRSAPRQNGAN 559

Db 1778 ALNQAIDQINHADKTAQVQNQNSIDAQNIISKIPATTVKATALQOIQNIATNKINLIKAN 1837

Qy 560 N-----LIFMGSKVK---TLGEMAFI-----SNRLHLD-----LSEQ 589

Db 1838 NEATDEEQNAAIQVQEKELIKAKQOIAGAVTNADVAYLLHDKGRKEIREIEPIVINKKATAR 1897

Qy 590 KOLTEIPVQAFSDNALKEVILLPASIKTRREAFKKNHLKOLEVASALSASHIAFNALDDNDG 649

Db 1898 EQLTTL-----FND---KQAEANVOATVEE--RNSILAQLQ---NIYDTRAGIQDQDRS 1945

Qy 650 DEQPDNKVVVKTHNSYALADGE-HFIVDPDKLSSTIVDLEKILKJIEGLDYSTLRQTQ 708

Db 1946 NAOVMDKTATL---NLQTIHLDVHPKPKDAEKTINDDLARVTHLVQ--NYRKVSDRNK 1999

Qy 709 TQFRDMTTAGK-----ALLSKSNLRQGEKQKFLQEAQPFGLGRVD----- 747

Db 2000 ADALKAITALKLQMDDEBLKTARTWADVADVKRFNVALGDIETAVITEKENSLLRIDNTAQ 2059

Qy 748 ----LDKAIKAAE-----KALVTKKATKNGKOLLERS-----INKAVLAYNNSA 786

Db 2060 QTYAKFKAIATPEQLAKVKKALIDQYVADGNRMVDSDATLNDIKKDTQLIIDELIAIKLPA 2119

Qy 787 -----IKKAN---VKRLEKELDLLTGLVEGKGPLAQATMWQGVYLL 824

Db 2120 EVIKASPKVGPAPKVCPTPIKEDKQEVKVVKELP-NTGSEMDLPLKELALITGAALL 2178

RESULT 43

US-10-732-923-3331

; Sequence 3331, Application US/10732923

; Publication No. US20050108791A1

; GENERAL INFORMATION:

; APPLICANT: Edgerton, Michael D

; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES

; FILE REFERENCE: 38-15(52796)C

; CURRENT APPLICATION NUMBER: US/10/732,923

; CURRENT FILING DATE: 2003-12-10

; PRIOR APPLICATION NUMBER: 10/310,154

; PRIOR FILING DATE: 2002-12-04

; NUMBER OF SEQ ID NOS: 24149

; SEQ ID NO 3331

; LENGTH: 1790

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-10-732-923-3331

Query Match 3.9%; Score 198.5; DB 17; Length 1790;

Best Local Similarity 21.1%; Pred. No. 0.0052;

Matches 249; Conservative 173; Mismatches 425; Indels 333; Gaps 59;

Qy 25 FSLVKEPILKQTOA--SSSISGADYAESGSKLKNINETSQPVDDTVD----- 71

Db 434 FDLQDFLLKQVQLCNNSTNNVGNNAKENGSGSKSDKSDKDDTGDGKGTGTEGSPKAN 493

Qy 72 LF-----GDKRTTPEKI-----KONLAKGPREQELKAVTENTSEKQITSG 112

Db 494 LFEVLLNYDAELNLNPPKLPFTTIDIFMFFQODH----KYSEELREITRN-----VTTG 543

Qy 113 SQLQSKESISLNTKVP---STS-----NWEICDPITKGNITLVGLSKS 152

Db 544 NDLE-DEEPLKAIQITISELLTSLTAADIRIPIISYLTFLIYWLFGDFKATNDFL--SDKS 600

Qy 153 GVEKLSQTHLVLPSSQAAD-----GTQLIQVA---SFAFTPKKTAIAEY-TSRAGEN- 201

Db 601 VIKSLLSFSYQI---QDEDVTIKCLVTMLLGVAEYFSSKSPFKPKSYFFFITKLGNKON 657

Qy 202 --GISQLDQVD--GKEIINEGEVFNLSYLV-----LKKVTIPTGYKHIQDAPV----- 244

Db 658 YASRIKQFKDQSYSKYKMDNEDSILTELDGTGLPKVYFSTYFIQLFNENIYRIRITALSH 717

Qy 245 --DNKNIAEAVNLPE-----SLETISDYFAFAHLAKQIDLPDNLKAIGEL 286

Db 718 DPDEBEPISKISFEVEKQLQKQCTKKGKITSQGTETESTHENLTKELIALTNEHKELDEK 777

Qy 287 AFDN-----QITKLSLIPQLMRLAERAFKSNHIKTIIEFRGNSLKV 328

Db 778 YQILNSSHSLKENFISILETELKKNVRDSDLEMTQLRDVLETKDKENQTALEVKYKSTIHK- 836
QY 329 IGEASFDNDLSQMLPDGLKEIESEAFGNPDHNNRVVLTWTSKGNPGLATENTY 388
Db 837 -----QEDSIK--TLEKLETLISOK---KKAEDGIN-----KMGKDLFALSREMOA 878
QY 389 VNPD-KSLWQESPEIDY-----TKWLEEDFTYQKNSVTGFSNKGKQVK-RNKNLEIPKQ 441
Db 879 VEENCKNLQKXKSNVNHQKETSLEKEDIAKITEIKAI-NENLEEMKIQCNNLSKEKE 937
QY 442 HNGVTITEIGD-----NAFRNVDQNTLRKYDLEEVKLPSTIR- 480
Db 938 HISKELVEYKSRFQSHDNLVAKLTEKLSANNYKDMAENESLIK-AVEESKNESIQL 996
QY 481 -----KIGAPAFQSNLKK-----SFEAS-----DDLEEIKEGAFMNNRIETLE- 518
Db 997 SNLQNKIDMSQEKENFOIERGSEIEKNIQLKTTISDLEQTKEEIIKSDSSKDEYESOI 1056
QY 519 --LKDKLVITGDA-AFHINHIYAIVLPESVQIEGRSAFRQGANNL-IFMGSKVKTGEM 574
Db 1057 SLUKEKLETATTANDENVNKISLTKTREETLEAELAAAYK-NLKNLETKLETSEKALKEV 1115
QY 575 AFLSNRLEHLDLSEQKQLETPVQAFSDNALKEVL-----LPASLTKTIRBEAFKKNHLK 628
Db 1116 KENEHLKEEKIQLEKEATEKQQLNSLRANLESLEKEHEDLAAQLKKEVEQIANKERQY 1175
QY 629 QLEVASALSHIAFNALDNDGDEQPDNVVTHNSYALADGEHFIV-----LEGEVKAMKSTSEOSNLKK 1222
Db 1176 NEEISQ-----LNDEITSTQENESIKKKND-----LEGEVKAMKSTSEOSNLKK 1222
QY 677 -DPDKLSSTIVDLEK-----ILKLEGLDYSTL-----ROTTQTPQDRMTT 716
Db 1223 SEIDALNLQIKELKKNETNEASLESIKSEVETVKIKELQDECNFKKEVSELEDKLK 1282
QY 717 AGKALLSKNLQKQKQKFLQ-----EAQFFLGRV-DLDKAIKAEKALVTKKAT--- 765
Db 1283 ASEDKNSKYLEQKSEKIKELDAKTTELKIQLEKITNLSKAKESELSRLKKTSS 1342
QY 766 --KNG-QLLERSINKAVLAYNNSAIKANVKELEKELDLTLGLVSGKPLAQATVQGVY 822
Db 1343 ERKNAEQLEKUNE--IQIKNOA-----FEKERKLN--EGSSTITO-----1381
QY 823 LLKTLPLPEYVIGLVNVPDKSGKL-----IYALDMSDTIGRGQDAYGNPILANVDEDE 877
Db 1382 -----EYSEKINTLEDELIRLQENELAKAIDNTRSELEKVSLSNDELLEKQN- 1431
QY 878 GYHALAVATLADYEGLDITILNSKLSQTSI-----ROVPTAAYHRAGIFQAIQNAAB 932
Db 1432 -----TIKSLOD-EILSVKDKITRNDKLLSTERNKRDLESLEKQ-----LRAAQESKAK 1481
QY 933 ABQLLPKPTHTSEKSSSESA---NSKD---RGLQSNPKTN 967
Db 1482 VEEGLUK-----LEEESSEKAELEKSKEMMKLESTIESN 1517

RESULT 44
US-10-369-493-1586
; Sequence 1586, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1586
; LENGTH: 1790
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1586

Query Match 3.9%; Score 197.5; DB 15; Length 1790;
Best Local Similarity 21.0%; Pred No. 0.006;
Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

QY 25 FSLVKEPILKQPA---SSSISGADYAESGSKJLKNETSQVDDTVTD----- 71
Db 434 FDLQDFLLKQVLCNNSTNNVGNKENGSGNSDKSDSDKDTGDKGTVEGSGFKAN 493
QY 72 LP-----SDKRTTPEKI-----KDNLAGPREQELKAVTENTSEKQITSG 112
Db 494 LFEVLNVDALNLPFKLFTTDFIMFFFOQDH---KYSBELREITRN-----VTTG 543
QY 113 SOLEQSKESLSLNTKTPV---STS-----NWEICDFTKGMTLVGLSKS 152
Db 544 NDLE-DEEPLKAIQIISLTTISLTAADIRIPISVLTLYWLFQGFKATNDFL--SDKS 600
QY 153 GVEKLSQTDHLVLPQAAD-----GTOLIOVA---SFAFTPDKKTAIAEY-TSRAGEN- 201
Db 601 VIKSLLSFSYQI---QDEDTVTKCLVMTLLGVAYEFSKESPPRKEYPEFITKTLGKN 657
QY 202 --GEISQLDVD--CKEIIINEGEV-----FNSYLLKKVT-----IPTGYKH 237
Db 658 YASRIKQPKQDSYFSKVDNMNEDSILTPELDEGLPKVYFSTYFIQLFNEINIRIATLSH 717
QY 238 IQQDAFVNKNIAEYNLPE-----SLETISDYAFALHAKLQDLPDLNLAIGEL 286
Db 718 DPDEPINKISFEVEKLEQROCTKJGITSQTTESTHENTLTKLIALTWEHKLDEK 777
QY 287 AFFDN-----QITKGLSLPRQLMRLAERAFKSNHINTIFRGNLSKV 328
Db 778 YOILNSSHSLKENFISILETELKKNVRDSDLEMTQLRDVLETKDKENQTALEVKYKSTIHK- 836
QY 329 IGEASFDNDLSQMLPDGLKEIESEAFGNPDHNNRVVLTWTSKGNPGLATENTY 388
Db 837 -----QEDSIK--TLEKLETLISOK---KKAEDGIN-----KMGKDLFALSREMOA 878
QY 389 VNPD-KSLWQESPEIDY-----TKWLEEDFTYQKNSVTGFSNKGKQVK-RNKNLEIPKQ 441
Db 879 VEENCKNLQKXKSNVNHQKETSLEKEDIAKITEIKAI-NENLEEMKIQCNNLSKEKE 937
QY 442 HNGVTITEIGD-----NAFRNVDQNTLRKYDLEEVKLPSTIR- 480
Db 938 HISKELVEYKSRFQSHDNLVAKLTEKLSANNYKDMAENESLIK-AVEESKNESIQL 996
QY 481 -----KIGAPAFQSNLKK-----SFEAS-----DDLEEIKEGAFMNNRIETLE- 518
Db 997 SNLQNKIDMSQEKENFOIERGSEIEKNIQLKTTISDLEQTKEEIIKSDSSKDEYESOI 1056
QY 519 --LKDKLVITGDA-AFHINHIYAIVLPESVQIEGRSAFRQGANNL-IFMGSKVKTGEM 574
Db 1057 SLUKEKLETATTANDENVNKISLTKTREETLEAELAAAYK-NLKNLETKLETSEKALKEV 1115
QY 575 AFLSNRLEHLDLSEQKQLETPVQAFSDNALKEVL-----LPASLTKTIRBEAFKKNHLK 628
Db 1116 KENEHLKEEKIQLEKEATEKQQLNSLRANLESLEKEHEDLAAQLKKEVEQIANKERQY 1175
QY 629 QLEVASALSHIAFNALDNDGDEQPDNVVTHNSYALADGEHFIV-----LEGEVKAMKSTSEOSNLKK 1222
Db 1176 NEEISQ-----LNDEITSTQENESIKKKND-----LEGEVKAMKSTSEOSNLKK 1222
QY 677 -DPDKLSSTIVDLEK-----ILKLEGLDYSTL-----ROTTQTPQDRMTT 716
Db 1223 SEIDALNLQIKELKKNETNEASLESIKSEVETVKIKELQDECNFKKEVSELEDKLK 1282
QY 717 AGKALLSKNLQKQKQKFLQ-----EAQFFLGRV-DLDKAIKAEKALVTKKAT--- 765

Db 1283 ASKDKNKYELQKSEKIKEELDAKTTELKIQLEKIITNLSKAKESSELSRLKTSSE 1342
Qy 766 --KNG-QLLERSINKAVLAYNSAIKKANVRLEKELDLTLGLVEGKPLAQATWQGVY 822
Db 1343 ERKNAEQLKQKNE--IQKNQA-----PEKERKLLN-----EGSSTITQ----- 1381
Qy 823 LLKTPPLPEYVIGLVYFGSGKL-----IYALDMSDTIGEGOKDAYGNPILNVEDNE 877
Db 1382 -----EYSEKINTLEDELIRLQENELKAEIDNTRSELEKVSLSNDELLEEKQ- 1431
Qy 878 GYHALAVATLADYEGDLTIKILNSKLSQTSI-----RQVPTAAYHRAGIFQAIQNAAE 932
Db 1432 -----TIKSLQD-EILSYKDKITRNDKLSIERDNKEDLSLKEQ-----LRAAQESKAK 1481
Qy 933 ABQLLPKPGTHSEKSSSSESA---NSKD--RGLQSNPKTN 967
Db 1482 VEEGLKK---LEBESSKEAELEKSKEMMKLESTIESN 1517

RESULT 45
US-10-732-923-3330
; Sequence 3330, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3330
; LENGTH: 1790
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-732-923-3330

Query Match 3.9%; Score 197.5; DB 17; Length 1790;
Best Local Similarity 21.0%; Pred. NO. 0.006;
Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

Qy 25 FSLVKEPIKQTOA---SSSISGADYAESGSKLKNINSGPVDVTVD----- 71
Db 434 FDLQRDFLLKQVLCCNNTNNVGDNAKENGSGNSDKESDSKDVTGDKDGYEGSFKAN 493
Qy 72 LF-----SDKRTTPEKI-----KDNLAGPREQELKAVTENTSEKQITSG 112
Db 494 LFEVLNLYDAELNLPKLFPTTDFIMFFQDQD---KYSEELREITRN-----VTIG 543
Qy 113 SOLQESKESLSINKTVP---STS-----NWEICDFITKGNLTVLGLSKS 152
Db 544 NDLE-DEEPLKAIQITSELTITSLTAADIRIPISYLTFLIYWLFGDFKATNDFL--SDKS 600
Qy 153 GVEKLSQTDHLVLPQAD-----GTQLIQA-----SFAPTPKKTAIAEY-TSAGEN- 201
Db 601 VIKLSLSPSYQI---QDESDVTIKCLVTMLGVAYEFSSKESFPFRKEFFETITKLGKN 657
Qy 202 --GETSOLDVD--GKEITNEGEV-----FNSYLLKKVT-----IPTGYKH 237
Db 658 YASRIKQPKKUSYFKVMDNEDSILTPDELGLPKVYFSTYFIQLFNIYRIRITALSH 717
Qy 238 IQQDAFVDMKNIAVNLPE-----SLETISDYAFAHALKQIDLDPNKAIGEL 286
Db 718 DPDEEPINKISPEEVEKLQROCTKLKGETISLTQETESTHENLTKLIATNHEKELDEK 777
Qy 287 APFDN-----QITGKLSLPRQLMRLAERAFKSNHIKTIEFRGNSLKV 328
Db 778 YQLNSSHSLKENFSIILETELKYNVDSLDEMTQLRDVLETKDKENQTALLEYKSTINK- 836
Qy 329 IGAEAFQNDLSQLMPOGLEKIESEAFNGPDHYNRVVLMWTKSGKNPSPGLATENTY 388
Db 837 -----QEDSIK--TLEKGLEITLSQK---KKAEDGIN-----KMGKOLPALSREMQA 878

Qy 389 VNPD-KSLWQESPEIDY-----TKWLEEDFTYQKNSVTGFSNKGLOKVK-RNKNLEIPKQ 441
Db 879 VEENCKNLQKOKSNVNHQKETSLSLAKEDIAAKITEIKAI-NENLEEMKIQCNLSSEKE 937
Qy 442 HNGVTITEIGD-----NAFRNVDFQNKTLRKLYDLEEVKLPSTIR- 480
Db 938 HISKELVEYKSRFQSHDNLVAKUTEKLSLANNYKQMAENESLIK-AVESKNESIQ 996
Qy 481 -----KITGAFAPQSNLUK---SFEAS-----DDLEIKEGAFMNNRIETLE----- 518
Db 997 SNLQNKIDMSQKQENFQIERGSIETKNIQKKTISDLQETKBERIISKSPSSKDEYESQI 1056
Qy 519 --LKQKLVITIGDA-AFHINHYIIVLPESVQIEGRSAFRONGANNL-IFMGSKVKTIGEM 574
Db 1057 SLLKBELETATTANDENVNKISELTKTREBLEAELAAAYK-NLKNELETKLTSETSEKALKEV 1115
Qy 575 AFLSNRLEHLDLSEKQLTPIPVQAFSDNALKEVL-----LPASLKTITREAFKQNHK 628
Db 1116 KENEEHLKEEKIQLEKEATETKQOLNSLRANLSLEKEHEHDLAQLKYYEIQANKEROY 1175
Qy 629 QLEVASALSHIAFNALDNDGDEQFDMKNVVKTHNSYALADGEHFIV-----RQTTQTOFRDMTT 716
Db 1223 SEIDALMLQIKELKCKKQNETNEASILLESIKSVESSTVVKIKELQDECNFKEKEVSELDKJK 1282
Qy 717 AGKALLSKSNLRQEKOKFLQ-----EAQFFLGRV-DLDKAIKAEKALVTKKAT--- 765
Db 1283 ASEDKNKYELQKESKIKEELDAKTTELKIQLEKIITNLSKAKESSELSRLKTSSE 1342
Qy 766 --KNG-QLLERSINKAVLAYNSAIKKANVRLEKELDLTLGLVEGKPLAQATWQGVY 822
Db 1343 ERKNAEQLKQKNE--IQKNQA-----PEKERKLLN-----EGSSTITQ----- 1381
Qy 823 LLKTPPLPEYVIGLVYFGSGKL-----IYALDMSDTIGEGOKDAYGNPILNVEDNE 877
Db 1382 -----EYSEKINTLEDELIRLQENELKAEIDNTRSELEKVSLSNDELLEEKQ- 1431
Qy 878 GYHALAVATLADYEGDLTIKILNSKLSQTSI-----RQVPTAAYHRAGIFQAIQNAAE 932
Db 1432 -----TIKSLQD-EILSYKDKITRNDKLSIERDNKEDLSLKEQ-----LRAAQESKAK 1481
Qy 933 ABQLLPKPGTHSEKSSSSESA---NSKD--RGLQSNPKTN 967
Db 1482 VEEGLKK---LEBESSKEAELEKSKEMMKLESTIESN 1517

RESULT 46
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match          3.9%; Score 197.5; DB 9; Length 2368;
Best Local Similarity 19.8%; Pred. No. 0.0088;
Matches 242; Conservative 165; Mismatches 459; Indels 359; Gaps 59;

QY 19 THNOEVSFLVKPILKQTOASSISGADYABSSG-----KSKLKINETSQVDDTVTD 71
Db 850 TQNAVSNATNEVEADAA-----VEAQOGLHDIQVVKSKOEVDATKSKVLDKINA 903

QY 72 LFSDKRTTPEKID-NLAKGPREQELKAVTENTESEKQITSGSQLESKESLSLNTKTVPS 130
Db 904 IQTOAKVKPAADTEVENAYNTRKQEIQNSNASTTEEKQ-AAYTELDTKKQEARTNLDAAN 962

QY 131 TSNWEICDPIIT-KGNTLVGLS-----KSGVVKLS----- 158
Db 963 TNS-----DVTTKANGGIAINOVOAATTKSKDAKAEIAQKASERKTAIEAMNDSTTEEQ 1018

QY 159 -----OTDHLVLPQOA-----ADGQLIQVASF-AFTPD-----KTAI-----A 192
Db 1019 AAKDKVDQAVVTANADIDNNAANTVDNNAKTNEATIAAITPDANVPKPAQAIADKVOA 1078

QY 193 EYTSRAGENGESQLDNDGKEIINEGEVFNYSLLKKVITPTGKHHGQDAFVD---NKN 249
Db 1079 QETAIDANNATTEKAAKQVQ-----TEKTTADTAIDGAHTNAVEAAKNAEI 1129

QY 250 AEVNPESLETSDYAFALHAKQIDLPNLKAIGEL-----AFPDNOITKGLSLP 300
Db 1130 AKIEAIQPAATTKDKNAQAIATKANERKTAIAQTQDITABEIAAANANVDNAV 1183

QY 301 ROLMLAERAFKSNHKTIEFRGNSLKVIGESFQDNDLSQLMLPDGLEKIESEAFGN- 359
Db 1184 -QANNIEAANSQNDV-----DOAKTTGEASIDQ-----VTPTVKKKATAVTDAXNN 1229

QY 360 ---PGDDHNNRVLMWTKSGKN-----PSGLATENTYVNPDKSLWQSPEDIDYTKWLEE 410
Db 1230 ITAATDD---NGVDYAKADAGKNSIQSTQATAVKSNKNDVDQAVTTQQAIDNTTGAT- 1285

QY 411 DFTYQKNSVTGSPNGKLOKVKRNKNLEIPKQNGVITIEIGNAFRV-----DFQNTKL 465
Db 1286 ---TEEKNAKGL-----VLKAKEKAYQDILNAQTNDVTQIKQAVADVQGITADTTIKDV 1339

QY 466 RKVDL-----BEVKLPSTI-----RKIGAFAPQSN-NLKSPEASDLEIEKEG 507
Db 1340 AKDELATKAREKALIAQTADATTEKEQANQOQVDAELTQGNQNIENASQSIDDVNTAKDN 1399

QY 508 AFM-----NNRIETL-ELKOKLVITIGDAAAPHINHIYAIVLPESVQIEG--RS 551
Db 1400 AIGADIPQASTDVKTNARAELLTEMQKITELNNNETNE-----EKGNIDIGPVRA 1452

QY 552 AFRQNGANNLIIPMGSKVTKLGEMAFLSNRLEHLDISEQKQLTEIPVQAFSDNALKEVLLP 611
Db 1453 AY-EGLGNNI---NAATTG-----DVTTKADTAVQVKVQLHANPVKK---P 1492

QY 612 ASLKTIREFAP--KKNHLKQL-----EVASALSHIAFNALDNDGDQFQDNK 656
Db 1493 AG-KTALDQAAADKTKTQIBQTPNASQOEINDAKQEVDTLNOAKTN-IDQSTDEYVDNA 1550
```

```
QY 657 V-----VVKT--HHNSYALADGEHF-----IVDPD-----KLSSTIVDL-EKILKLEIG 697
Db 1551 VKEGKAKINAVKTFSEYKKDALAKTEAAAYNAKAVTEADNSNASTSSSIEAEAKQKLAELKOT 1610

QY 698 LDYSTLRQTTOTQFRDMTTAGKALLSKNSLKQGEKQKFLQEAQFPL-----GRVDLDKATA 753
Db 1611 ADQNVNQATSKDDIEVQIHNDLNDINDYTIPTGKESATTDILYAYADQKKNNISADTNAT 1670

QY 754 KAEKALVTKKATKNSQLLERSINKAV-----LAYNNSAI-----KKAN----- 791
Db 1671 QDEKQQAIAKQVQNVQTALESINNGVDNGVDVDDALTOGKAALDAIQVDAIVKPKANQVTD 1730

QY 792 -----VKRLEKELDLLTG-----LVEGKGFLAQATMVQGVYLLKTLPLPLEY 833
Db 1731 AKABETKESIDQSDQLTAEKTEALAMIKQITDQAKQGITDATTTAEEKAKAQ----- 1784

QY 834 YIGL-----NVYFDKSGK-----LIYALDM-----SDTIGEGQKDAYGNPILNVDBD 875
Db 1785 --GLEAFDNIQIDSTEKQKAIELEFALTALDOIEAGVNVVDADATTE-EKEAFTNAL----- 1835

QY 876 NEGHALAVATLADY-EGLDIKTILNSKLSQTSIRQVPTAAHYHRAGI-----FQAI 926
Db 1836 -EDILSKATEDISDQTTNABEATVKNLSALEQKQRIINPVVKGNALAEIAREVVNKKIEII 1894

QY 927 QNAAAEAE-----QLLPKPGTTHSEKS-----SSSESAN 954
Db 1895 KNADADASAKBIARTDLGRYFDRPADLKDQTKTQNTNEVAELQNVITPAIEAIVPQNDPNAN 1954

QY 955 SKDRGLQSNPKTNRGHSAIILPRTG 979
Db 1955 DTNSGSDNNDATANSNANATPENTG 1979
```

RESULT 47

```
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12389
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12389
```

Query Match 3.9%; Score 197.5; DB 9; Length 2368;
Best Local Similarity 19.8%; Pred. No. 0.0088;
Matches 242; Conservative 165; Mismatches 459; Indels 359; Gaps 59;

QY 19 THNQVFLVKEPILKQTOQASSISGADYABSSG-----KSKLKINETSPPVDDTVD 71
DB 850 TONATVSNATNEEVAEADA-----VEAQKQGLHDIQVVKSKQEVADTKSVLKDINA 903
QY 72 LPSDKRTTPEKID-NLAKGPREQLKAVTENTSEKQITSGSLQBSKESLSLNTKTVPS 130
DB 904 IQTQAKVPAADTEVENAYNTRKQEIQNSASTTEBKQ-AAYTELDTKKQEARNLDAAN 962
QY 131 TSNWEICDPIT-KGNTLVGLS-----KSGVEKLS----- 158
DB 963 TNS-----DVTAKONGIAINQVQAATTKSDAKAIEAKASERKTAIEAMNDSTTEBQ 1018
QY 159 -----QTDHLVLPQQA-----ADGTQLIQVASF-AFTPD-----KKTAI-----A 192
DB 1019 AAKOKVQAVVTANADIDNAANTDNDNAKTNEATIAAITPDANVKPTAKQAIADKVOA 1078
QY 193 EYTSRAGENGESQLDVGKEIINEGEVFNVSLLKKVTIPTGYKHIGQDAFVD---NKNI 249
DB 1079 QETAIDANNAGTTEEKAQAQVQ-----TEKTTADTAIDGAHTNAEVEAKNAEI 1129
QY 250 AEVNLPSLETISDYAFALHAKQIDLPNLKAIGEL-----AFPDNQITGKLSLP 300
DB 1130 AKIEAIQPATTKONAKQAIATKANERKTAIAQTODITABEIAAANAVNDNAV----- 1183
QY 301 RQLMLAERAFKSNHIKTIERNGLSKVIGBASFOQNDLSQMLPDGLEKIESEAFNG- 359
DB 1184 -QANNIEAANSQNDV-----DQAKTTGEASIDQ-----VTPVVKKATAVTDACKN 1229
QY 360 ---PODDHNNRVVLWTSKGN-----PSGLATENTYVNPDKSLWQSPSEIDYTKLBE 410
DB 1230 ITAATDD---NGVDATKADGKNSIQSTQPATVAKSNAKNDVDQAVTTQQAIDNTTGAT- 1285
QY 411 DFTYQKNSVTGFSNGKLVKKNLEIPKQHNGVTITEIGNAFNV-----DFQNKTL 465
DB 1286 ---TEEKNAKOL-----VLKAKEKAYQDILNAGTNDVTQIQDQAVDQGGTAUTTIKDV 1339
QY 466 RKYDL-----EVLKLPSTI-----RKIGAFAPQSN-NLKSPEASDDLEEKEG 507
DB 1340 AKDELATKAREQKALIAQTADATTEKEQANQOVDDELTOGNQNLNENASIDVNTAKDN 1399
QY 508 AFM-----NNRIETL-ELKOKLVITIGDAAFHINHIYAIVIPESVQETG---RS 551
DB 1400 AIQAIDPIQASTDVKTNARAEILLTEMQNKITEILANNETTNE-----EKGNDIGPVRA 1452
QY 552 APRQNGANNLIPMGSKVKTIGEMAFLSNRLEHLDLSEOKQLTEIPVOAFPSNALKEVLLP 611
DB 1453 AY-REGLNNI-----NAAITTG-----DVTAKOTAVQKVQQLHANPVKK---P 1492
QY 612 ASLATTIREAF--KGNHLKQL-----EVASALSHIAFNALDDNDGDPOFNK 656
DB 1493 AG-KTALQQAADKKTQIEQTPNASQOEINDAKQEVDTLNOAKTN-IDQSTDEYVDNA 1550
QY 657 V-----VVKT--HNSYALADGEHF-----IVDPD-----KLSSTIVDL-EKILKLEEG 697
DB 1551 VKEGKAKINAVKTFSEYKDALAKTEAAYNAKVTEADNSNASTSEIAEAKQKLAELKQT 1610
QY 698 LDYSTLROTTQTPRDMTTAGKALLSKNLRGKQKFLQEAQFPL-----GRVLDLKAIA 753
DB 1611 ADQNVNQTASKDDIEVQIHNDLNDINTYIPTGKESATTDUYAYADQKKNISADTNAT 1670
QY 754 KAEKALVTKCATKNGQLLERSINKAV-----LAYNNSAI-----KKAN----- 791
DB 1671 QDEKQQAIKQVDNQVQTALESINNGVDNGVDVDDALTQGKAAIDAIDQVDAATVKPKANQVID 1730
QY 792 -----VKLEKELDLTG-----LVEBGKPLAQNTWQGVVLLKTPLPPEY 833
DB 1731 AKAEETKESIQDSQDLTAEEKTEALAMIKQITDQAKQGITDATTTAEVEKAKAQ----- 1784
QY 834 YIGL-----NVYPDKSGK-----LIYALDM-----SDTIGEGQKDAYGNPILNVDB 875

DB 1785 --GLEAFDNTQIDSTEKQKAEIELETALDQIEAGVNDADATTE-EKEAFPTNAL----- 1835
QY 876 NEGYHALAVATLADY-EGLDIKTILNSKLSQLTSSIRQVPTAAYHRAGI-----FQAI 926
DB 1836 -EDILSKATEDISQDTTNAETATVKSALQKQRIQNPVVKKNALBAIREVNNKQIEII 1894
QY 927 QNAAABAE-----QLLPKPGTHSEKS-----SSSSSAN 954
DB 1895 KNADADASAKIEARTDLGRYPDFADKDKTQTNTVEAELQNVVTIPAIEAIVPQNDPNAN 1954
QY 955 SKDRGLQSNPKTNGRHSAILPRTG 979
DB 1955 DTNSGSDNDATANSNANATPENTG 1979

RESULT 48
US-10-369-493-22285
; Sequence 22285, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22285
; LENGTH: 1875
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22285

Query Match 3.9%; Score 196.5; DB 15; Length 1875;
Best Local Similarity 20.8%; Pred. No. 0.0073;
Matches 242; Conservative 164; Mismatches 378; Indels 379; Gaps 59;

QY 34 KQTQASSI-SGADYASSSGSKUKINET-----SGPVDDTVDTL-----FSDKRT 78
DB 199 RKTQELTLQSNNDWLEKELRSK---NEQYLSYRQKTDKVIDIRNELNRLRNDPQMBT 255
QY 79 TPEKIK---DNLAGPREQ--ELKAVTENTSEKQITSG-----SOLEQSKE 120
DB 256 NNDVLKQKNNELSKLOEKLLEIKGLSDLSLNSKQEFSAEMSLKQRLVDLLESQNLAVKE 315
QY 121 SLS-----LN--KTVPSTSNWEICDFTKGNLTVLGSKSGVEKLSQTDHVLPLSQA----- 169
DB 316 ELNSIRELNTAKVIADDSKKQ---TPENEDLLKELQLTKEKLAQCEKCLRLSSITDEA 371
QY 170 -ADGTQLIQVASFAPDPKTAIAEYTSRAGENGESIQOLDVDGKE---IIN----- 216
DB 372 DEDNENLSAKSSDFIPLKQLIKERRTKHQLQIETFFIVELEHKVPFIINSFKERTDML 431
QY 217 EGEVFNVSLLKKVTIPTGYKHIGODAFVDNKNIAEAVNLPSLETISDYAFALHAKQIDL 276
DB 432 ENELNNAALL-----LEHTSNE---KNAKVLELNKQO-----KLVEC 466
QY 277 PDNLKAIGELAFFNQITGKLSLPRQLMRLAERAFKSNHIK-----TIEF----- 321
DB 467 ENDLQTLTK-----QRDLRCQIQYLLITNSVNSDSKGLRKEBIOFIQINIMQEDD 517
QY 322 ----RGNLSKVIGBASFOQNDLSQIM-----LPGCLE-----KIESE 354
DB 518 STITESDSQKVTVRELVEFKNIIOLEKNAELLKVVRLNADLKLESKEKKSKQSLQKIESE 577
QY 355 AFTGNPGDDHNNRVVLWTSKGNPSGLATENTYVNPDKSLWQSPSEIDYTKLWEEDFTY 414

Db 578 --TVNEAKE-----AIITLKSEK-----MDLESRIEELQELBELKTS 613
 Qy 415 QKNSVTGPNKGLQKVRNK-NLEIPKOHNGVTITEIGNAFRNVDFONKTLRKVDLEEV 473
 Db 614 VPDNADSYNVTIKQLTETKRDLESQVQDLOTRISQITRESTENSLLNKEIQ--DLYDS 671
 Qy 474 KLPSTIRKIG-----AFAFQSNLNKSFASDDLEIEKEGAFMNNRI-----514
 Db 672 KSDISI-KLGKESKRILAEERFKLLSNTLDLTKAEND-QLRKRDPYQLNTLIKQDSKTH 729
 Qy 515 ----ETLEKDKLVITIGDAFHI--NHIYAVLPESV-QEIGRSAPFRQNGANNLIFMGSK 567
 Db 730 ETLENYVSCSKLSIVETELLNKEQKLRVHLEKNLQELNKLSPK--DSLRIWVTO 786
 Qy 568 VKTLGEMAFNLRLHLDLSEQKLTETIPVQAFSDNALKEVLLPASLTIRREAPKKNHL 627
 Db 787 LQTLQKER--EDLLEETRSCKQKIDEL-----EDALSE-----LK--KETSQKHII 830
 Qy 628 KOLEVASALSHIAFNALDNDGDEQDNKV--VVKTHNSYALADGEHFI VDPDKLSSTI 685
 Db 831 KOLEE-----DNNSEIWEYQNKIBALKKDYESVITSVDSKQ--TDIEKLYQKV 876
 Qy 686 VDLEK-----ILKUIEG-----LDYSTLROTT-----707
 Db 877 KSLEKEIEBDKIRLHTYNNVMDETINDDSLRKELEKSKINLTDAYSQIKEYKDYETTSQS 936
 Qy 708 --QT-----QFRDMTTAGKALL-SKSNLRQGEKQKFLQEAQFFL-GRVDLDKATAKAEK 757
 Db 937 LQOTNSKLDESFPDFTNQIKNLTDEKTSLE--DKISLLKEQMFNLNLDLQKGMEXEK 994
 Qy 758 ALVTKATKNGQLLERSINKAVLANNNAIK-----KANVKRLEKELDLTLGLVEGKGP 811
 Db 995 ADFKKRI-----SILQNNKEVEAVKSEYESKLSKI QNDLDOQT-----1033
 Qy 812 LAQATWQGVYLLKTLPLPEYIYGLNVYFDKSGKLIYALDMSDTIGE--GOKDAYGNPI 869
 Db 1034 IYANTAQNNY-----EQLQKHADVSKTISELRQLHTYKQV 1071
 Qy 870 --LNVDED-----NEGYHALAVATLADYEGLDIKTLNLSKLSQTSIRQVPTAAVHR 919
 Db 1072 KTLNLSRQLENALKENESKWSQKESL--LEQLDLS--NSRIEDLSSQNKL--LYDQ 1123
 Qy 920 AGIFOAIQNAAEAEQLPKPGTHS-----EKSSSES 952
 Db 1124 IQIYTA--ADKEVNSTNGPGLNLIITLRERDILDTKVTVAERDAKMLRQKLSMDV 1180
 Qy 953 ANSKDRGLQSNPKTNRGRHSAIL 975
 Db 1181 ELQDARTKLDNSRVEKENHSSII 1203

RESULT 49

US-10-732-923-3334
 ; Sequence 3334, Application US/10732923
 ; Publication No. US20050108791A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
 ; FILE REFERENCE: 38-15(52796)C
 ; CURRENT APPLICATION NUMBER: US/10/732,923
 ; CURRENT FILING DATE: 2003-12-10
 ; PRIOR APPLICATION NUMBER: 10/310,154
 ; PRIOR FILING DATE: 2002-12-04
 ; NUMBER OF SEQ ID NOS: 24149
 ; SEQ ID NO 3334
 ; LENGTH: 1875
 ; TYPE: PRN
 ; ORGANISM: Saccharomyces cerevisiae
 US-10-732-923-3334

Query Match 3.9%; Score 196.5; DB 17; Length 1875;
 Best Local Similarity 20.8%; Pred. No. 0.0073;

Matches 242; Conservative 164; Mismatches 378; Indels 379; Gaps 59;

Qy 34 KOTQASSSI-SGADYABESSGSKLKINET-----SGPVDVDTVDL-----FSDKRT 78
 Db 199 RKTQELTLLQSNNDWLEKELRSK---NEQYLSYRQKTDKVIDIRNELNRLRNDFOKERT 255
 Qy 79 TPEKIK---DNLAKGPREQ--ELKAVTENTSEKQITSG-----SOLEQSKKE 120
 Db 256 NNDVLKQNNELSKSLQELKLEIKGLSDLSNLSKEQFSAEMSLKQALVDLLESQNAVKE 315
 Qy 121 SLS-----LN--KTVDPSTSNWEICDPTKGNITVLGLSKSGVEKLSQTDHLVLPQA-- 169
 Db 316 ELNSIRELNTAKVIADDSKKQ---TPENEDLLKELQLTKEKLAQCEKELRSLSSITDEA 371
 Qy 170 -ADGTQLLIQVASFAPTPDKTAIAEYTSRACENGESIQLDVDGKE--IIN-----216
 Db 372 DEDNENLSAKSSDPIFLUKQLIKERRTKHEHLQNOETFI VELEHKVPIIINSFKERTDML 431
 Qy 217 EGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFALHALKQIDL 276
 Db 432 ENELNNAALL-----LEHTSNE--KNAKVKELNAKQ-----KLVEK 466
 Qy 277 PDNLKAIQELAPFNOITGKLSLPQLMRLAERAPKSNHIK-----TIEP-----321
 Db 467 ENDLQTLTK-----QRILDLCRQIQYLLITNSVNSDKGPLRKEBIOFTINQMOSDD 517
 Qy 322 ---RGNLSKLVIGASFOQNDLSQLM-----LPDGLE-----KIRSE 354
 Db 518 STITESDQKVVTERLVEFKNIIQOEKNAELLKVVRLADKLESKEKSKQSLOKISE 577
 Qy 355 AFTGPGGDHYNRRVWLTKSGKNPSGLATENTYVNDKSLWQESPEIDYTKWLEEDFTY 414
 Db 578 --TVNEAKE-----AIITLKSEK-----MDLESRIEELQELBELKTS 613
 Qy 415 QKNSVTGPNKGLQKVRNK-NLEIPKOHNGVTITEIGNAFRNVDFONKTLRKVDLEEV 473
 Db 614 VPDNADSYNVTIKQLTETKRDLESQVQDLOTRISQITRESTENSLLNKEIQ--DLYDS 671
 Qy 474 KLPSTIRKIG-----AFAFQSNLNKSFASDDLEIEKEGAFMNNRI-----514
 Db 672 KSDISI-KLGKESKRILAEERFKLLSNTLDLTKAEND-QLRKRDPYQLNTLIKQDSKTH 729
 Qy 515 ----ETLEKDKLVITIGDAFHI--NHIYAVLPESV-QEIGRSAPFRQNGANNLIFMGSK 567
 Db 730 ETLENYVSCSKLSIVETELLNKEQKLRVHLEKNLQELNKLSPK--DSLRIWVTO 786
 Qy 568 VKTLGEMAFNLRLHLDLSEQKLTETIPVQAFSDNALKEVLLPASLTIRREAPKKNHL 627
 Db 787 LQTLQKER--EDLLEETRSCKQKIDEL-----EDALSE-----LK--KETSQKHII 830
 Qy 628 KOLEVASALSHIAFNALDNDGDEQDNKV--VVKTHNSYALADGEHFI VDPDKLSSTI 685
 Db 831 KOLEE-----DNNSEIWEYQNKIBALKKDYESVITSVDSKQ--TDIEKLYQKV 876
 Qy 686 VDLEK-----ILKUIEG-----LDYSTLROTT-----707
 Db 877 KSLEKEIEBDKIRLHTYNNVMDETINDDSLRKELEKSKINLTDAYSQIKEYKDYETTSQS 936
 Qy 708 --QT-----QFRDMTTAGKALL-SKSNLRQGEKQKFLQEAQFFL-GRVDLDKATAKAEK 757
 Db 937 LQOTNSKLDESFPDFTNQIKNLTDEKTSLE--DKISLLKEQMFNLNLDLQKGMEXEK 994
 Qy 758 ALVTKATKNGQLLERSINKAVLANNNAIK-----KANVKRLEKELDLTLGLVEGKGP 811
 Db 995 ADFKKRI-----SILQNNKEVEAVKSEYESKLSKI QNDLDOQT-----1033
 Qy 812 LAQATWQGVYLLKTLPLPEYIYGLNVYFDKSGKLIYALDMSDTIGE--GOKDAYGNPI 869
 Db 1034 IYANTAQNNY-----EQLQKHADVSKTISELRQLHTYKQV 1071
 Qy 870 --LNVDED-----NEGYHALAVATLADYEGLDIKTLNLSKLSQTSIRQVPTAAVHR 919
 Db 1072 KTLNLSRQLENALKENESKWSQKESL--LEQLDLS--NSRIEDLSSQNKL--LYDQ 1123

```
QY 920 AGIFOAIONAABAEABQLPKPGTHS-----EKSSES 952
Db 1124 IQIYTA---ADKEVNNSTNGPGLNLTIRREDRILDTKVTVAERDAKMLRQISLMDV 1180
QY 953 ANSKDRGLQSNPKTNRGHSAIL 975
Db 1181 ELQDARTKLDNSRVEKENHSSII 1203

RESULT 50
US-10-732-923-3335
; Sequence 3335, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgeton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3335
; LENGTH: 1875
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae
US-10-732-923-3335

Query Match 3.9%; Score 196.5; DB 17; Length 1875;
Best Local Similarity 20.8%; Pred. No. 0.0073;
Matches 242; Conservative 164; Mismatches 378; Indels 379; Gaps 59;

QY 34 KQTQASSRI-SGADYAESGSKLKINET-----SGPVDVDTVDL-----FSDKRT 78
Db 199 RKTQELTLQSNNDWLEKLSK---NEQYLSYRQKTDKVLDIRNELNRLNDFQMERT 255
QY 79 TPBKIK---DNLAKGPRQ--ELKAVTENTSEKQITSG-----SOLEQSK 120
Db 256 NNDVLKQKNELSKLOKLEIEIKGLSDLSNKEQEPFSAEMLKQRLVDLLESQNLAYKE 315
QY 121 SLS-----LN--KTVPSTGNWEICDFTKGNLTVLGSKGVEKLSQTDHLVLPQSA----- 169
Db 316 ELNSIRELNTAKVIADDSKKQ-----TPNEDLLKELQITKEKLAQCEKELRLSITDEA 371
QY 170 -ADGTQLIQVASFATPPDKKTAIABYTSRAGENGESIQDVGKE---IIN----- 216
Db 372 DEDNENLSAKSSDFFLKQLIKERRTKEHLQNIETPIVELEHKVPIINSFKERTDML 431
QY 217 EGEVNSVLLKKVTIPTGKYHGQDAFVDNKNIAEVNLPESLETISDYAPAHILAKQIDL 276
Db 432 ENELNNAALL-----LEHTSNB---KNAKVKELNKNQ-----KLVEC 466
QY 277 PDLNKAIGELAFFDNIQTKLSLPLQLMLAERAFKSNHK-----TIEF----- 321
Db 467 ENDLQTLFK-----QRDLCHQIQYLLITNSVNSDKGPKRKEBIOPTIOMQBEDD 517
QY 322 ----RGNSLKVIGESAFQNDLSQLM-----LPDGLE-----KIBSE 354
Db 518 STITESDSQVVTRELVFKNIIQLQKNAELLKVVRNLADKLESKEKSKQSLQKIBSE 577
QY 355 AFTGNPGDDHNNRVVLTWKSNKPSGLATENTYNNPKSLWQESPEIDYTNWLEEDFTY 414
Db 578 --TVNEAKE-----AIIITKSEK-----MDLESRIEELQKLEELKETS 613
QY 415 QKNSVTGFSKGLQKVRNK-NLEIPKQHNGVTITEIGNAFPRNVDFONKTLRKVDLEEV 473
Db 614 VPNEADASYNTVIKQLTETKRDLESQVQDLQTRISQITRESTENMSLNKSIQ--DJYDS 671
QY 474 KLPSTIRKIG-----APAFQSNLKSFEASDDLBEIKEGAFMNNRI----- 514
Db 672 KSDISI-KLGKESRILAEERFKLLSNTLDLTKAEND-QLRKRPDYLQNTILKQDSKTH 729
```

```
QY 515 ----ETLELKDKLVTIGDAAFHI--NHIIYAIVLPESV-OEIGRSAFRQNGANNLIFMGSK 567
Db 730 ETLNVEYCSKSLKSIVETELLNLKEBQKLRVHLEKNLKOELNKLSPK---DSLRIMVTQ 786
QY 568 VKTIGEMAFSLNRLHLDLSEQKOLTEIPVQAFSDNALKEVLLPASLKTIRREAFKKNHL 627
Db 787 LQTLQKER--EDLLEETKSCQKIDEL-----EDALSE-----LK--KETSQKDDHI 830
QY 628 KQLEVASALSIAFNALDNDGDQFQNKV--VVKTHNSYALADGBHFIVDPDKLSSTI 685
Db 831 KQLEE-----DNNSEIWEYQNKIEALKDYVESVITSVDSKQ--TDIEKLOQKV 876
QY 686 VDLQK-----ILKLIQEG-----LDYSTLRQTT--- 707
Db 877 KSEKEIEEDKIRLHTYNNVMDETINDDSLRLKEKSKINLTDAYSQIKYKDLYETTSQS 936
QY 708 --QT-----QFRDMTTAGKALL--SKSNLRQGEKQKFLQEAQFPL-GRVDLDKAIKAEK 757
Db 937 LQQTNSKLDSEFKDFTNQIKNLTDEKTSLE--DKISILLKEQMFNLNNELDLQKKGMEKEK 994
QY 758 ALVTKATKQGLLERSINKAVLAYNNSAIK-----KANVKRLEKELDLTLGLVEGKGP 811
Db 995 ADFPKRI-----SILQNNKEVEAVKSEYESKLSKIQNDLDOQT----- 1033
QY 812 LAQATWQGVYLLKTPLPPEYIYGLNVYFDPKSGKLIYALDMSDTIGE--GQDAYGNPI 869
Db 1034 -IYANTQNNY-----EQLQKHADVSKTIISLRBQLHTYKQV 1071
QY 870 --LNVDED-----NEGYHALAVATLADYEGDIDKTLNLSKLSQTSIRQVPTAAVHR 919
Db 1072 KTLNLSRQLENALKENEKSWSSQKESL--LEQLDLS---NSRIEDLSSQNKL---LYDQ 1123
QY 920 AGIFOAIONAABAEABQLPKPGTHS-----EKSSES 952
Db 1124 IQIYTA---ADKEVNNSTNGPGLNLTIRREDRILDTKVTVAERDAKMLRQISLMDV 1180
QY 953 ANSKDRGLQSNPKTNRGHSAIL 975
Db 1181 ELQDARTKLDNSRVEKENHSSII 1203

RESULT 51
US-10-831-070-6
; Sequence 6, Application US/10831070
; Publication No. US20050112612A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Alterman, Eric
; APPLICANT: Cano, Raul J.
; APPLICANT: Hamrick, Alice
; TITLE OF INVENTION: Lactobacillus Acidophilus Nucleic Acid
; TITLE OF INVENTION: Sequences Encoding Cell Surface Homologues and Uses
; TITLE OF INVENTION: Therefore
; FILE REFERENCE: 5051.690
; CURRENT APPLICATION NUMBER: US/10/831,070
; CURRENT FILING DATE: 2004-04-23
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 2539
; TYPE: PRF
; ORGANISM: Lactobacillus acidophilus
US-10-831-070-6

Query Match 3.9%; Score 196; DB 17; Length 2539;
Best Local Similarity 19.1%; Pred. No. 0.012;
Matches 237; Conservative 178; Mismatches 440; Indels 388; Gaps 59;

QY 27 LVKPEIILKQTOASSISGADYAE---SSGKSKLKINETSGP-----VDDTFT 70
Db 507 LIKEATDAANNAKAIDKATTAIDAQDEBGTNNINNVTPVPSLEDAKKAATKAVDALD 566
```


Db 772 NLELHISKLEHENVSELSEFISGLESQLTLYLANEKELSMQMDERSLITNLKDE-LEQVE 830
 QY 432 RNK-NLEIPKQNGVITTEIGNAFNVDVFNKTLR-----KYDLEVKLPS 477
 Db 831 AQKVELKQMDERSLITNLKDE-LEQVEAQKVELKQMDERSLITNLKDELEQVE--- 886
 QY 478 TIRKIGAFAPQSNLKSFP-----EASDDLEEIKEG-----AFNNRIE-----TLBK 520
 Db 887 -AQKV---ELKENQLESHRRLESEVQSDSEALRSNAKLQATVDHVVECKSLQTLTADLK 942
 QY 521 DKLVITGDAAPHINHIYAIVLPESVQETGRSAFRQNGANNLIFMGSKVKTLGEM-AFLSN 579
 Db 943 KQKLEV-----HGYASHLEQLEQSKKTM-----DFCKTLESLEAKLSS 982
 QY 580 RUEHLDLSEKQKLTIP---VQAFSDNAKVELLPASLKTIREAPKQNLKQL-----EVAS 634
 Db 983 LQEDISLKEQSLSELENIQFQSHKEHERIDRVHLLNKIEKTEKVELHFPFLEREVIS 1042
 QY 635 ALSHTAPNALDNDGDEQFDNKVVVTKHNSVALADGEHFIVDPDKLSSTIVDLE----- 689
 Db 1043 LTAQL-----SSTEERESSTLDTIREVSTLRADKAKLEANLEDVNAQMIHYESQLED 1095
 QY 690 -----KILKLEGL-----DYSTLRQTTQTFQDMTTAGKALLS---KSNLR 728
 Db 1096 LRESKTKIKOLVDSLNASQNEEMLTDDVNMRSIEARSNEDNLRKTLCELELKSXS 1155
 QY 729 QOEKQKF-----LQEAQFFLGRVD-LDKAIA-----K 754
 Db 1156 DYEQQIITEEISVLKIQVHKIAGLQDEVLTQLQSSLEAKFEKGLQGLIQSLSECEBLK 1215
 QY 755 ARKALVTKK-----ATQOQLLERSINKAVLAYNNSAIKANVRELEKELDLLTGL 805
 Db 1216 AQKGLMTDKVSCMQDTLNAANEKGQIEISAQTKVLMGDEPPVKETSDVLEAKLSLSI 1275
 QY 806 VEGKPLAQATWQGVYLLKTPLPPEYVIGLVNVPFDSKGLIYALDMSDITIGEQKDAY 865
 Db 1276 IRG-----ANSEYQKIKYLSQEE---NEDLTRNQLMEKE-----LDKTS----- 1313
 QY 866 GNPILNVBDENGYHALAVATLADYEGLDIKTILNSKLSQLTISIRQVPTAAYHRAGIFQA 925
 Db 1314 ----QNKDENTNKQVSLQDEVLMQLSSLDLQSSLEAKFEKGLGL----- 1350
 QY 926 IQNAAAEARQLPKPGTHSEK-SSSESANSKDRGLQ 961
 Db 1351 LQSLSECEELKAQKGLMTDKVSCMQDTLNAANEKGQ 1387

RESULT 53

US-10-282-122A-60961
 / Sequence 60961, Application US/10282122A
 / Publication No. US20040029129A1
 / GENERAL INFORMATION:
 / APPLICANT: Wang, Liangu
 / APPLICANT: Zamudio, Carlos
 / APPLICANT: Malone, Cheryl
 / APPLICANT: Haselbeck, Robert
 / APPLICANT: Ohlsen, Kari
 / APPLICANT: Zyskind, Judith
 / APPLICANT: Wall, Daniel
 / APPLICANT: Trawick, John
 / APPLICANT: Carr, Grant
 / APPLICANT: Yamamoto, Robert
 / APPLICANT: Forsyth, R.
 / APPLICANT: Xu, H.
 / TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 / FILE REFERENCE: ELITRA 034A
 / CURRENT APPLICATION NUMBER: US/10/282,122A
 / CURRENT FILING DATE: 2003-02-20
 / PRIOR APPLICATION NUMBER: 60/191,078
 / PRIOR FILING DATE: 2000-03-21
 / PRIOR APPLICATION NUMBER: 60/206,848
 / PRIOR FILING DATE: 2000-05-23
 / PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 60961
 ; LENGTH: 1778
 ; TYPE: PRT
 ; ORGANISM: Listeria monocytogenes
 US-10-282-122A-60961
 Query Match 3.8%; Score 195.5; DB 15; Length 1778;
 Best Local Similarity 20.2%; Pred. No. 0.0078;
 Matches 209; Conservative 137; Mismatches 330; Indels 361; Gaps 50;
 QY 44 GADYAESGSKKINETSQVDDVTDFLFDKRTT--PEKIKD-NLAKGPREQELKAVT 100
 Db 27 GADTTVSDEAVTATRADTSATEGIESGTGDDTAEPEKAEKASKEETEEKAKT 86
 QY 101 ENTESEKQI---TSGSQLEQSKESLSLNKTVPTSTSNMEICDPTFKGNTLV---GLSKSGV 154
 Db 87 EEPASNKTEINTDKSQLKQT---SLKAAVPAGSTV-----NSLPPDDNLAK--- 130
 QY 155 EKLQTDHLVPSQAADGTQLIQVASFAFTPKTKATAEYTSRAGEINGEISQLDVDGKEI 214
 Db 131 -KLAV---ITGNAATAGNESVDSALL-----AISQLDL----- 161
 QY 215 INEGEVFNYSLLKKVTPPTGYKHIGQDAFVFNKNAEVLNLPESLETISDYAFAPAHALKQI 274
 Db 162 -SGETGND-----PTDISNIEGLQYLE-NLTSNLSEN--NISDLA---PLK-- 201
 QY 275 DLPNLKAIGELAFPDNQITGKLSLPRQLMRLAERAPKSNHIKTIE--FRGNSLKVIGEA 332
 Db 202 ---DLNVLNLSLSSNRTLVNLSGVEDLVNLQELNVSAN--KALEDISQVASLPVLKEI 255
 QY 333 SPQNDLSQLMPLDGLKIESEAFGNPGDDHYNRVVVLTKSKGNPSGLATENTYVNP 392
 Db 256 SAQGCNTKLEL-----KNPAGAVL----- 275
 QY 393 KSLWQESPEIDYTKWLEEDPTYQKNSVTGFSNKGLOKVRKNKLEIP-----KOHNG 444
 Db 276 -----PEL-----ETFYQENDLTNLTS--LAKLPKLNLYIKGNASLSKLETLNG 319
 QY 445 VT-----ITRIGD-----NAPRNV-DFQNKTLRKVDL 470
 Db 320 ATKQLIDASNCTDLETIGDISGLSELEMIQLSGCSKLKEITSLKNLPNLVNITADSCAI 379
 QY 471 BEVKLPSTIRKIGAFQSN-NLKSPASDDELEIK-----EGAFNN-----RIETL 517
 Db 380 EDLGTIANLPKLTQLVLSDNENLTNITAITDLPQKTLTLDGCCITSIGTLNDLNPKEKL 439
 QY 518 ELKDKLVITGDAAPHINHIYAIVLPESVQETGRSAFRQNGANNLIFMG--SKVTKLGBMA 575
 Db 440 DLKE-----NQITSI---SEITDLPRLSYLDVSVNNLTITGDKLKLPLEHLN 484
 QY 576 FLSNRLHLDLSEKQKLTIEPVQAF---SDNALKEVLLPASLKTIREAFKKNHLKQLEV 632
 Db 485 VSSNR-----LSDVSTLTNFPPLNVIINNNVIRTVGQMTLPSLKEFYAQNNSIDISM 539
 QY 633 ASALSHI-----AFNALDDNDGEQFDNKVVVTKHNSVALADGEHFIVDPDKLSSTIV-- 686

Db 540 IHMPNLRKVDASNNLTNIG--TFDN-----LPKLSLDVHSNRTSTSVTH 585
Qy 687 DLEKILKLEGLDYSTLROTTQTPROMTGTAGKALLSKNLRQGEKQKFLQEAQFPGLGRV 746
Db 586 DLPSL-----ETFWAQTNLNIG-----TMDNLPD-----LTVNLSFNRI 622
Qy 747 -----DLDAKAIAEALVTKKATNGOLLERSINKAVLAYNSAIKKANVKRLE 796
Db 623 PS LAPIGDLPNLETLLVSDNNLSYLSRLGTMWG-----VPKLRILDLQ 664
Qy 797 KELDLTLGLVEGKPLAQATWQGVYLLKTPPLPYIYGLNVYFK-SG-----KLIYA 850
Db 665 NNYLNTYGTENGLSSLDLT-----NLTELNRNNVYIDISGLSLTSLRIY- 711
Qy 851 LDMSDTIGEGQKDAYGNPILNVDEDN-EGVHALA-----VATLADYEG 892
Db 712 -----LNDNKIEDISALSNLTNLQELTLENKNIENISALSDLEN 752
Qy 893 LDKITLNSKLSQTSIROVPTAAYHRAGIFQA-----IQNAAAEAEQLPKP 940
Db 753 LNKLVVSKNKIIDISPVANN-----VNRGAIVTASNQTYLPTVLSYQSSFTIDNPVWYD 808
Qy 941 GTHSEKSSSESANSKD 957
Db 809 GTLLAPSSIGNSGNKYD 825

RESULT 54
US-10-724-972A-5942
; Sequence 5942, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; PRIOR FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 5942
; LENGTH: 3696
; TYPE: PRN
; ORGANISM: S.epidermidis
US-10-724-972A-5942

Query Match 3.8%; Score 195; DB 18; Length 3696;
Best Local Similarity 18.8%; Pred. No. 0.024;
Matches 220; Conservative 164; Mismatches 478; Indels 310; Gaps 45;

Qy 10 LUTTVSVVTHNQEVFLSKPELIPKQTOASSISGADYAESGKGLKINETSFGVDDTV 69
Db 1082 IHTNVNVQKPKQARQALLAKTNEKQSAINSDEGT--IEEKQATQSLDANKLADEQI 1139
Qy 70 TDLFSDKR-----TTPEKIKONLAK-----GPRQELKAVTENTSEKQ 108
Db 1140 TQAASNQNVNALTIGISNISKIQTNTFTKKQARDQVQNKQFQKEAELNSTPHATQDEKQ 1199
Qy 109 ITSQSQESKESLSLNTKVPSTSNWEICDFITKG-----143
Db 1200 -DALTELTAQKET-ALNDINQATQNVQVDTALTSGIQTQNVNVRKQKEAKTTINDIV 1257
Qy 144 -----NTLVLSKSGVEKLSQTDHLVLPQ-----AADGTQLIQV 178

Db 1258 OQHQSQIQNDDATTTEEKEVANLVNASQNV--ISKIDNATTNNQIDGIVSDGRQIN- 1314
Qy 179 ASFAFTP-----DKKTAIAEYTSRAGE-----NGEISQLDVGKEII- 215
Db 1315 ---ALTPDTSIKRNNKNDIDIKAAKKIKIQINDATDEEIQEANKRKEAKIEAKONIQ 1371
Qy 216 -----NEGEVFNYSLLKKVTIPTGYKHIOQDAFVNDK-----NIAEVLNPESL 258
Db 1372 RNSTRDQVNEAKTNGINKIENITPATTVKSEARQA-VQNKANEQINHIOQTDPATNEEQ 1430
Qy 259 ETISDYAFALHAKQIDIDPDNLKALGELAPFNQITGKLSPLQML--BLASRAFKSNHI 316
Db 1431 EAINRVS-AELARVQAQNAEHTTQGVKTIKDDATISLSRINAQVVEKESARNAIQKAT 1489
Qy 317 KTFIEPRGNSLVIGFASQDNDLSQLMLPDGLEKIESEAFNPGDGDHNNRVLVWTKGS 376
Db 1490 QOTQFINNDNATDEKEVANNLVITATQKSLDNINSLS-----SNNDEVNAKVAG 1540
Qy 377 KN-----PSGLATENTYVNDKSLQWESPEIDYTKWLEEDFTYQKNSVTGFSNKLQKV 430
Db 1541 INEIANVLPAVAKSKAKKIDOKLAQIQNIQ-----THQ-TATTEEKEAAIQLA 1590
Qy 431 KKNKY--LEIPKQNGVTIITEIGDANAFNVDFQNKTLRKYDLEEVKLPSTTRKGA--- 484
Db 1591 NQKSNEARTAIQNEHSNNGVAQAKSNGIHEI-----ELVMPDAHKKSDAKOS 1637
Qy 485 ---FAFOSNNLKSFEASDLEEEIKEGAFMNNRIETLELKOKLVITIGDAAPHINIIYIV 540
Db 1638 IDNKYNEQSNTINT--TPDATDEEKOKAL-----DKLKTAKONGY--NKVDQQAQ 1682
Qy 541 LPESVQEIGRSAFRQNGANNLIFMGSKVKVTLGEMAFPLSNRLHDLSEKQKLTETPQVAF 600
Db 1683 TNQOVSDAKTEAI--DTITNI--QANVAKPSARVELDSKPEDL--KQINATP---- 1730
Qy 601 SDNALKEVLLPASLTIREEPKONHLKOLEVASALSHIAPNALDDNDGDEQFNKVVVK 660
Db 1731 --NATEEE-----KQDATQRLNGKRDEVKNLIN-----QDRRDNEV--- 1764
Qy 661 THNSYALADGEGHFIVDPDKLSSITVDI-----EKILKLEGL 698
Db 1765 EOHKNIGLOELETIHANPTRKSDALQELQTKPISQTELINNNKDATNEBEKDEAKRLLETS 1824
Qy 699 DYSTLRQTTQTO-----PRDMTTAGKALLSKSNLRQGEKQKFLQEAQFPGLGR- 745
Db 1825 KNKITINNOAQTNNOVDNADKONGNVEIATIPATITKDTAKTADKKAEQOVTTINGNN 1884
Qy 746 --VLDLKAIAK--AEKALVTKKATKNGOLLERSINKA-----VLAYNN---SAIKKANVR 794
Db 1885 DATBEKAEARKLVEKAKIEAKSNITNSDTEREVNGAKTNGLEKINNIQPSQTQKTNAQ 1944
Qy 795 -----LEKELDLTT-----GLVEGKPLAQATWQGVYLLKTPPL 830
Db 1945 EINDKAAQSLQIQTNPDPATBEKQKQATNRVNAQAQAIQINNNNAHSTQEVNESKTNSTA 2004
Qy 831 PEYITGLNVYFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDENEGVEGHALAVATLAD- 889
Db 2005 TIKSVQPNVI-----KKPTAINSLTQEAANNQKTLIGNDGNATDDKEAKQKLVQKLINEQ 2059
Qy 890 ----YEGLDIKITLNSKLSQTSIROVPTAAYHRAGIFQAIQNA--AEAEQLPKPGTH 943
Db 2060 IQKIHSTQDQNVQNVKAQAITAKLINANAHKRQDAINILTNLAESKKSDIRAMQDATT 2119
Qy 944 SEKSSSSSANSKORGLQSNPKTNRGRHSAIL 975
Db 2120 BEKNTAIQSID--DTLAQRNNINGANTNALV 2149

RESULT 55
US-09-815-242-11501
; Sequence 11501, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert


```

; LENGTH: 2186
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match      3.8%; Score 194; DB 9; Length 2186;
Best Local Similarity 19.1%; Pred. No. 0.013;
Matches 216; Conservative 183; Mismatches 387; Indels 342; Gaps 58;

QY 20 HNOEVFSLVKEPILKQTOASSISGADYAESGSKLJINETSQVDDTVDFLPSD--- 75
DB 1208 YNAKLAELINATPDATNDEKNAALINTLNQDRQQAIESIKQANTNAEVDDAAVAENNDIV 1267

QY 76 DNEGYHALAVATLADYEGDIDIKTLINSLKSLQTSIRQVPTAAYHRAGIFQAIQNAAEAE 934
DB 2092 D-----ATLN-----DIKQHTQFIVDEILAIAK-LPAEA-----TKVSPKEI 2126

QY 935 QLLPKPGTHSEKSSSESANGSKDRQSQNPKTNRGRHSAILPRTGSKG 982
DB 2127 QPAPKVCPIKKEETHESRKEKE-----LPNTGSEG 2158

RESULT 57
US-10-470-048B-336
; Sequence 336, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; TITLE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
; FILE REFERENCE: SONN:0315US
; CURRENT APPLICATION NUMBER: US/10/470,048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 336
; LENGTH: 2186
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-470-048B-336

Query Match      3.8%; Score 194; DB 17; Length 2186;
Best Local Similarity 19.1%; Pred. No. 0.013;
Matches 216; Conservative 183; Mismatches 387; Indels 342; Gaps 58;

QY 20 HNOEVFSLVKEPILKQTOASSISGADYAESGSKLJINETSQVDDTVDFLPSD--- 75
DB 1208 YNAKLAELINATPDATNDEKNAALINTLNQDRQQAIESIKQANTNAEVDDAAVAENNDIV 1267

QY 76 -----KETTPEKIKONLAKGPREQELKAVTENTSEKQITSGSOLEQSKE----- 125
DB 1268 QVDVVKQQAARDKITAIEVAK--RIEAVKQTPNATDEEQ--AAVQINQLKDAQAINQIN 1324

QY 126 KT---VPSTSNWEI-----CDFITKGNLTVGLSKGVEKLSQTDHLVLPSSQAADGTOL 175
DB 1325 QTNQDQVDTTQAVNAIDNVEAEVVIKPKAIDIEKAVEKQQAIDNSL---DSTDNEK- 1380

QY 176 IQVASFAFPDKKTAIAEYTSRAGENGEISQLDVGKEIINEGEVFNFSYLLKKVT--IPT 233
DB 1381 -EVAQALAKEKALAA-IDQAQTSQVNOQAATNGVSAIK-----IIQPETKVKPA 1430

QY 234 GYKHIGQDAPVDNKNIAEVLNPE-----SLETISDYAPAHAL-----KQID 275
DB 1431 AREKINOKA---NELRAKINQDKEATEABERQVALDKINE--FVNQAMTDTITNRTNQOVD 1485

QY 276 LPDNLKAIAGELAFPDNQITGKLSLPROLMRLAER-----AFKSNHIKTIIEPRGNSLKV 328
DB 1486 -DTSQALDSIA-----LVTPDHIIVRAAARDAVKQYEAQKREIEQAEHATDEEQ 1535

QY 329 IGEASFQDND-----LSQLMLPDGLEKIESEAFNGPDGDHNNRVYLVTKSGKNPSGL 382
DB 1536 VALNQLANNEKRALQNIDQAIANNNDVKRVETNGIATLKG---VQPHIVI-----KPEAQ 1587

QY 383 ATENTYVNPDKSLWQESP-----EIDYTKWL-----BEDFTYQKNSVTGFSN--- 424
DB 1588 AIKASAENQVESI-KOTPHATVDELDEANQLISDTLQKQAQOEIENTNQDAAVTVDRNQTI 1646

QY 425 KGLQKVK-----RNKN-----LEIPKQNGVTIIEIGD--NAFRNVD 459
DB 1647 KAIEQIKPKVRKRAALDSIENNNKQLDAIRNTLDTTQDERDAIDTLNKIVNTIKNDI 1706

QY 460 FQNKTKRYDLEE-----VKL-----PSTIRKIGAFQSNLKSPEASD-DLEEI 504
DB 1707 AQNKTAEVDRITETDGNQIKVILPKVQVPAARQSVGVKABQNAL--IDQSLSTEE 1764

QY 505 XEDA-----FMNRIETLELKDVLITGDAAFINHIYAIVLP-----ESVQEIG 549
DB 1765 RLAAKHLVBEALNQAIQINHADKTAQVNOQSINAQNIISKIKPATTVKATALQIQNIA 1824

QY 550 RSNAPRQNGANN-----LIFMGSKVK---TLGEMAFI---SNRLBHL 585
DB 1825 TNKINILIKANNEATDEEQNIAIAVEKELIKAKQOQIASAVTNADYAVILLHDKNEIREIE 1884

QY 586 -----LSQKOLTEIPVQAFSDNALKEVLLPASLKTIRREAFKKNHLKOLEVASALSHI 639
DB 1885 PVNIRKASAREQTTL-----FND---KQAIENAIQATVER--RNSILAQLQ---NIYDT 1932

QY 640 AFNALDDNDGDEQFQDN--KVYVYKTHNSYALADGEHFTVDPDKLSSTITVDLEKILKIEG 697
DB 1933 AIGQIDQDRSNAQVKTASNLQTIHDL-----DVHPICKPDAEKTINDOLLARVALVQ- 1986

QY 698 LQYSTLRQTTQTFQFRDMTAGKALLSKSNLROGEKQFLQBAQFPLGRVLDKAIKAEK 757
DB 1987 -----NYRKVSNRNK-----ADALK 2001

QY 758 ALVTKKATKNGOLLERSIN---KAVLAYNNSAIKKANVRLKELDLTLGLVEGKGLAQ 814
DB 2002 AITALKQMBEELKARTNADVDVILKRFNVALSDIEAVITEKNSLLR--IDN---IAQ 2056

```


Db 2617 ETELQREMHM--AOKTAELOEEL--SGEKRLAGELQLLLEI--K 2657
Qy 764 ATKNGOLLERSINKAVLAYNNSAIKKA-----NVKRLKELDLTLGLVECKGPLAQATWQV 819
Db 2658 SSKD-QLKE-----LTLENSELKSLDMKMDQVEKGVREBEIAEYQLRLHEAEKKG 2709
Qy 820 GYVLLKTPPLPEYIYGLNVYFDK-----SGKLIYALDMSDTIG 858
Db 2710 QALLLDTN---KQYEVEIQTYREKLTSSKECLSSQKLEIDLKSSKEELNNSIKATTQIL 2766
Qy 859 EGOKDA-----YGNPILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQ 906
Db 2767 BELKTKMDNLKYVNLKKNENARAQGMKLLIKSCCKQLE--BEKEILOKELSOLAQAEK 2824
Qy 907 -----TSIRQVPTAAYHRAGIFQAIQNAAAAEQOLLPKPG-----THSEKSSSESAN 954
Db 2825 QKTGTWMDTKVDELTEIKE---LKETLEKTKAEDEYLDKYCSLLISHEKLEKAKEMLE 2881
Qy 955 SKDRGLQSNPKTNRGHSAIL 975
Db 2882 TQVAHLCSQSQKDSRGSPLL 2902

RESULT 62
US-10-751-736-99
; Sequence 99, Application US/10751736
; Publication No. US2004026230A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Martinez, Robert
; APPLICANT: Brown, Eugene
; APPLICANT: Liu, Wei
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING AND TREATING COLON
; FILE REFERENCE: AM100927 (031896-002000)
; CURRENT APPLICATION NUMBER: US/10751,736
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US Provisional Application 60/438,000
; PRIOR FILING DATE: 2003-01-06
; NUMBER OF SEQ ID NOS: 54873
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 99
; LENGTH: 3114
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-751-736-99

Query Match 3.8%; Score 192.5; DB 16; Length 3114;
Best Local Similarity 20.2%; Pred. No. 0.026;
Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

Qy 6 KTVALTTLTVSVVTHNQVPSLVKBPILKQTOASSISGADVAESSGSKLKNITSGPV 65
Db 1973 KTVALTDLSEKMKETQEL-----ESHQSECLHCIOVAEAEVKEKTELLQT---L 2019

Qy 66 DDTVTLDFDKRTTPKPI---KONLAKGPREQELKAVTENESKQITSGSQLEQSKES 121
Db 2020 SDDVSELLKDKTHLQKQSLQSLKCELENQIAQLNKEKELL-----VKRES 2072

Qy 122 LSLINKTVPTSTNWEICD-----FITKNTLVGLS-----KSGVEKLS---QTD 161
Db 2073 ESIQARL--SESDYEKLVNSKALEALVEKGEFALRLSSQTQEEVHQLRRGIEKLVRIEAD 2131

Qy 162 -----HLV--LPSQAADGTQLQVASFATPPKKTAIABYTSRAGNGEISOLDVDG--- 211
Db 2132 EKKQLHIAEKLERENDSL-----KDKVENLERELQWSEENQELVILDAENSKA 2182

Qy 212 -----KEIINEGEVENSYLKKVTIPTGYKHI-----GQ-----DAF 243
Db 2183 EVETLTKTQIEEMARSLUKVFE---LDLVTLRSEKENITKQIQKQGLSBLDKLLSFKSL 2239

Qy 244 VDNKNIAEVLNPESETTISDYAFALHALKQIDLPDNLKAIGELAFPDNQITGKLSLPRQL 303

Db 2240 LEEKEQABITQIEESKTAVEM-----LQNLKEL-----NEAVALQCDQEI 2281
Qy 304 MRLAERAFKSNHIKTIERGNLSLVIGEASPDNDLSQMLPDGLEKLESEFTNGPGDD 363
Db 2282 MKATEQSLLDPPTEEBEHLR-NSIEKL-RARLEADEKKQLCVLQQLKSESHHA-----D 2332
Qy 364 HYNRRV-----VLMTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417
Db 2333 LLKGRVENLERELEARTNQEHAAEAENS-----KGEVETLKAKTEGWTQ--- 2378
Qy 418 SVTGSNKGLO-----KVRKNKLEIPKQNGVTTITEIGNAFNPRVDFQNKTKRYDL 470
Db 2379 -----SLRGLELDVVVTIRSEKNTLNLQKEQERISELEIINSFENI--LQSEKEQEKVQM 2432
Qy 471 BEVKPLSTIRKIGAFQSNLKSFEASDDLEIEIK-EGAFMNNRIETLELKLVTIGDA 529
Db 2433 KEKS--STAMENLOTQLKELNERNVAALHNDQACAKQONJSSQVECELEKQAQLQGLD 2490
Qy 530 APHINHIYAIVLPEVSV---QBIGRSAFRQNGANNLIIPWGSKVKTILGEMAFLSNRLEHD 585
Db 2491 EAKNYY---IVLQSSVNGLIQEV-----EDGKQKL-----EKDEEISRLKNQIQ--- 2532
Qy 586 LSEKQKLTPIPVQAFSDNAL-KEVLLPASLKTIRBEAFKXNHLKQLEVASALSHTAFNAL 644
Db 2533 --DQQLVSKLSQVEGEHQLWKEQNL--ELRNLTVLEQKIQVLQSKNAS-----L 2579
Qy 645 DDNDGDEQDNKVVVYKTHNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLR 704
Db 2580 QD-----TLEVLQSSYKNLENELELTMDKMSF-----VEKVNOM-----TAK 2617
Qy 705 OT-TOTQFRDMTTAGKALLSKNSLNQGEKQKFLQEAQFFLGRVLDLDAKAKAEKALVTKK 763
Db 2618 ETELQREMHM--AOKTAELOEEL--SGEKRLAGELQLLLEI-----K 2658
Qy 764 ATKNGOLLERSINKAVLAYNNSAIKKA-----NVKRLKELDLTLGLVECKGPLAQATWQV 819
Db 2659 SSKD-QLKE-----LTLENSELKSLDMKMDQVEKGVREBEIAEYQLRLHEAEKKG 2710
Qy 820 GYVLLKTPPLPEYIYGLNVYFDK-----SGKLIYALDMSDTIG 858
Db 2711 QALLLDTN---KQYEVEIQTYREKLTSSKECLSSQKLEIDLKSSKEELNNSIKATTQIL 2767
Qy 859 EGOKDA-----YGNPILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQ 906
Db 2768 BELKTKMDNLKYVNLKKNENARAQGMKLLIKSCCKQLE--BEKEILOKELSOLAQAEK 2825
Qy 907 -----TSIRQVPTAAYHRAGIFQAIQNAAAAEQOLLPKPG-----THSEKSSSESAN 954
Db 2826 QKTGTWMDTKVDELTEIKE---LKETLEKTKAEDEYLDKYCSLLISHEKLEKAKEMLE 2882
Qy 955 SKDRGLQSNPKTNRGHSAIL 975
Db 2882 TQVAHLCSQSQKDSRGSPLL 2902

RESULT 63
US-10-661-809-3
; Sequence 3, Application US/10661809
; Publication No. US20040101919A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOCINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR
; TITLE OF INVENTION: POSITIVE BACTERIA
; FILE REFERENCE: P07741US01/BAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 1010

```
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
us-10-661-809-3

Query Match      3.8%; Score 191; DB 16; Length 1010;
Best Local Similarity 19.6%; Pred. No. 0.0067;
Matches 215; Conservative 159; Mismatches 391; Indels 334; Gaps 50;

QY 111 SGLSEQKESLSLNKTVPSNWEIC--DFTTK-----GNTVLGSLKSGVGEKLSQTDHL 163
Db 30 SGEKLNQALS VKLAKQSVTAKDVQITVKGFINKGVGDGNTTV-----DQQL 77

QY 164 VLPSQAA-----DGTQLIQVASFAFPDPKKTAAEY 195
Db 78 TIPANVAINEEKTTPSSLTLOWDVTEATSEVERDGTGVFGNIQTNTATDFGFSLSHT 137

QY 196 SR--AGENG--EISQ-----LDVDGKEIINEGEVFNYS-----LLKVT--IPTGY 235
Db 138 FVRVAVGKNGVSEWSEBPIKGTQDDPYKETINQVATSNLPEQPGAELKLTDKDLSTGM 197

QY 236 KHIGQDAFVDDNKNIAEVLNPSLETISYAFALHAK-----QID-----LPDNLKAIG 284
Db 198 -----HTWSTGIANPSDGNF--LSLKFDLGAEYQMDKIEYLPDRNAGNG 240

QY 285 EL-----APDNOITCKLSLPRQLMLAERAFKSNHIKTIEFRGN----- 324
Db 241 NILQOYRTSGDGNWTFSEBPIINWKQDALTKTIETKQAVRFVEMKVLKSVGNFGSGRE 300

QY 325 -----SLKVIGEASFQNDLSQLMLPDGLKIESEAFGTG--NPGD----- 362
Db 301 MLFYKQPGTEGILHGDINDGTIDENDAMSRYNTGLSVDSDF--FNGYVEKGLDNKNGVI 359

QY 363 DHYNRVVLWTKSG----- 380
Db 360 DAYDISYVLRQDGGIEIPDVEBEIAGGLSLAVVNBNGKDTLPGLDTLFILKQDGLKNIN 419

QY 381 GLAENTYVNPDKSIWQSSPEIDYTKWLEEDTYOKNS-----VTGFSNKLQKVKNK 434
Db 420 ALSTKMSDFSSFEVLGVPATNTQOMENYSKYRKHSDNENLYLVLSNQG-----NK 473

QY 435 NLEIPKHNGVITTEIGDNFANRVPFQNKTLAKYDLEEVKLPSTIRKIGAF----- 485
Db 474 QL-----LNG-----SMDLVPFKVKVKTTRVKRATTVEQPLQFDMSSQLLVG 516

QY 486 -AFQNNLKSFEASDDBEIKEGAFMNNRIETLEKOK-----LVTIGDAAPHINH----- 535
Db 517 QGFQATLSDFSVT-----VKPTLVLDKELLQALITLQOARVEKEYTPT 561

QY 536 --IYALVLPESVQEIGRSFRON-----GANNLIPMGSKVKTLGEMAFLSNRLEHLDLSEQ 589
Db 562 WAIFKPILDEAVAVLANEQATQDVSAAENLEKASQLEKRPDVA---NKAD--LEKAIQ 617

QY 590 KOLTEIPV--QAFSDNALK--EVLPLPASLKTIRBEAFKKNHLKQLEVASALSHIAFNALD 645
Db 618 EGLAKKPSDQGFETETKVLBSLAAQKVPQAEKVQEEIDQ----- 661

QY 646 DNDGDEQFNDKVVVTHNSYALAGEHFIVDPDKLSSTIVDLEKILKLIBGLDYLTURQ 705
Db 662 -----ATKTLREAIQLKEQVPVAVDKETLKEQIAQ--ARGRKPEGYQFT---K 705

QY 706 TTOTQFDMTTAGKALLSKNSLRQCEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKA- 764
Db 706 ETEKQLQEAIAQAEIAIVAKETATKEVSEALNALETAMAQLKEVPLVNDKQLEVVKRAQ 765

QY 765 ----TKNGOLLERS---INKAVLAYNNNAIKK--ANVKRLEKELDLLTGLVEGKGLAQ 815
Db 766 QVTPSEGHQFTASSIQLQELQKALLAAKNT--LKNPAANQKMWIDEAVELTSAIDG---LQEE 821

QY 816 THVQGVYLLKTPLEPIEY--IGLVYFDKSGKLIYALDMSDTTGBGQKQYGNPIINV 872
Db 822 VLVTDKKALEAMIAKAKIPKSPAGKEFTSESKARLTEAIDQAEGL--LADKNARQEQI--DI 879

QY 873 DEDNEGYHALAVATLADYEGLDIKTILNSKLSQLSIRQVPTAAVHRAG--IFQAIQNA 930
```

```
Db 380 AEKN-----VKTALD--SLE-EQVLTQDKTKKELLQKAEETLKPAGAKFTKASQDAL 929
QY 931 AE-----AEQLLPKPGTHSEKSSSESANSK-----DRGLQSNPKTNRGRHSAILPRTG- 979
Db 930 AEAIKQAKALVEDFNATQEAVDKCLSLQIAEAMABEPISSNSTGNNGHSTVSGTGV 989
QY 980 ---SKGSFVYIGLYTSTA 995
Db 990 TSQKGTATGTTTKTTTS 1008

RESULT 64
US-10-282-122A-71235
; Sequence 71235, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71235
; LENGTH: 3692
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-71235

Query Match      3.8%; Score 190.5; DB 15; Length 3692;
Best Local Similarity 19.9%; Pred. No. 0.044;
Matches 230; Conservative 184; Mismatches 399; Indels 343; Gaps 56;

QY 30 EPILKQTOASSISGADYAESGSKLKINETSGPVDVTDLPFSDKTTPEKI--KNL 87
Db 2696 ENILPATTVKDKAKADVNAKEKN--LQINS-----NDEATTEKLVASDNL 2741

QY 88 AKGPREQELKAVTE-----NTESEKQI-----TSGSQLEOS--KESLSL 124
Db 2742 -NHVETTNQAIEDAPDPTNQVNEKNKGIGTIRDIQPLVVKKPTAKSKIESAVEKKTEI 2800
```


; ORGANISM: Schizosaccharomyces pombe		US-10-732-923-3328	
Query Match		3.7%;	Score 189.5; DB 17; Length 1957;
Best Local Similarity		20.6%;	Pred. No. 0.021;
Matches		250;	Conservative 175; Mismatches 451; Indels 339; Gaps 59;
Qy	37	QASSISGADYAESGSKLKINETSGPVDVT	-----DLFSDKR 77
Db	4	QSSSGNTSLDBESASS--LVSSAASPFDSDLETPRPNISRASTGQAEDGDTSSQHE	61
Qy	78	TYPEKIDKNLAPRQELKAVNTESKQITSGSQLESLSLNTKVTSTNWEIC	137
Db	62	DSSEELKQEVRRHSLSIDAKLGSSEGSTASS-----ALPLTPRSPSNASW----	111
Qy	138	DFITKNTLVGLSKS---GVEKLSQVDHLVLPQAAADGTQLIQVASFAFTPKDKTAIAEY	194
Db	112	-LLVRG---GLLDSPLIDINSVTQKSNLL-----NELKQVRS-----KLALEH	151
Qy	195	TSRAGENGIS-QLVDGKEIINEGEVFN-----SYLLKKVTIPTGYKHIGQDAFVN	246
Db	152	-----ENGILSLQSSNKKDKNTSSVTLTSEEDVSYFQKKLT-----NMESNFA	198
Qy	247	KNAEVLNLPESLETISDYAFALHAKQID---LPDNLKAIGELAFPDNQITKLSLPRQL	303
Db	199	KOSEAYDLRQLLTVTE---KLDKKEKYEKIKEDVSSI-KASLAEEQASNK-SLRGQ	252
Qy	304	MLAERAFKSNH-IKTIEFRGNSL-----KVICE-----ASFOD	336
Db	253	ERLEKLLVSNKTVSTLROTENSLRAECTYQKLEKCAINEEDSKLLELKHNVANSYSD	312
Qy	337	NDLSQLMLPDGLEKIESEAFTPGDDHYNNRVLMTKSGK-----NPSG-----LATEN	386
Db	313	AIVHKDLIEDLSTRISEP-----DNLKSERDTLSIKNEKLEKLRNTIGSKDSRTSN	366
Qy	387	TYVNDPKSLWQSPEDIDYTKW-----LEEDFYQKNVSGFSGKLGQVKRNKULE	437
Db	367	SQLEEEVNLKESNRTIHSQLTDAESKLSFQENKSLSGSIDYQNNLSKDKMKVQVS	426
Qy	438	IPKQ-----HNGVTITIGDNAPRNVDPQNTLRKYDLEVKLPSTIRKIGAFQAFON	490
Db	427	SQLEEARSLAHATGKLAEI--NSER--DFQNKIKIDFEKIQDLRACLNS-----SSN	476
Qy	491	NLKSFEASDLEBEIKGAFMNNRIETLEKDKLVTTIGDAAFHI-----NHIAI	539
Db	477	ELKEKSALID---KKQQLNLRREQIKEQKVSESTQSSLOSQDLQILNEKKGHEVYES	532
Qy	540	VLPEVQETGRSAFRONGANNLIFMGSKVTGLG---EMAFNLRLHLDLSEKQKLTETP	596
Db	533	QLNELKQEL-----QTEISNGEHLSSQLSTLAAEKEAAVATNN-----ELSESKNLSQTL	582
Qy	597	VOAFSDNALKEVL-----LPASLKTIRE--EAFKQNH--LKOLEVASA-----	635
Db	583	CNAFQEKLAQVNLKENSQNFSSLDTSFKKLNESHQELNNHTTITQLDKDTSKLOQL	642
Qy	636	-LSHTAFNALDNDGDGEQDNKV-VVKTHNSVALADGEHFTVDPDKLSSTIVDL-EKIL	692
Db	643	QLERANFEQKSTLSDENNDRLTKLKBESKSLIKQOB---DVSLEKNITQLKEDLR	699
Qy	693	KLIEGLDYSTL-----RQTQTFQDMTTA-----GKALLSKSNLRQOE	731
Db	700	KSHEALRFKLEAKNLREVIDNLKGKHETLEAQRNDLHSSLSDAKNTNAILSELTKSSE	759
Qy	732	KQKFL-----QEA-----QPLGRVDLDKAIA-----KAEKA	758
Db	760	DVKRLTANVETITQDSKAMKQSTSLVNSYQISNLHYELRHHVNMQSNNTLSESK	819
Qy	759	LVT--KKATKNGQLLERSINKAVLAVNSAIIKANVRLKELDL-----	801
Db	820	LKTDENLTQONTLLIDNVQKLMHKHVNQESKVELKGVNGKLSLDLKNLRSNLVAISD	879
Qy	802	----LTGLVE---GKGPLAQ---ATMVQGVYLLKTPLP-----PEYIGLVNYYFDK-SGK	846

Db	880	NDQILTLAELSKNYDSLEQESAQLNSGLKSLAEKQLLHTENEELHRL-----DKLTGK	935
Qy	847	LIVALDMSDTTIGEGOKDAYGNPILNVDBDNEGYHALAVA-----TLADYBGL--DIKT	897
Db	936	LKIESKSDLGK-KLTARQEBISNLKEENMSQQAITSVKSKLDETISKSLLEADIEH	994
Qy	898	ILNSKLSQUTSIRQVPTAAYHR-----AGIFOAIQNAAAAEQQLLPKPTHSE	945
Db	995	LKN-KVSEVEVERNALLASNERLMDLKNNGENIASLQTEIEKKRAENDDLQSKLSVSS	1053
Qy	946	K-----SSSESANSKDRGLQSNPKT-----NRGHSAILPRTGSKGFFVGILOY	991
Db	1054	EYENLLLSSTQNKSLDKTNQKYEKNVQKLDKQDNVLEBELTSK-----YKLCGE	1109
Qy	992	TSVALLSLITAIKK 1006	
Db	1110	ENAIQKDELLALRKK 1124	
RESULT 68			
US-10-282-122A-70176			
; Sequence 70176, Application US/10282122A			
; Publication No. US20040029129A1			
; GENERAL INFORMATION:			
; APPLICANT: Wang, Liangsu			
; APPLICANT: Zamudio, Carlos			
; APPLICANT: Malone, Cheryl			
; APPLICANT: Haselbeck, Robert			
; APPLICANT: Ohlsen, Kari			
; APPLICANT: Zyskind, Judith			
; APPLICANT: Wall, Daniel			
; APPLICANT: Trawick, John			
; APPLICANT: Carr, Grant			
; APPLICANT: Yamamoto, Robert			
; APPLICANT: Forsyth, R.			
; APPLICANT: Xu, H.			
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms			
; FILE REFERENCE: ELITRA.034A			
; CURRENT APPLICATION NUMBER: US/10/282,122A			
; CURRENT FILING DATE: 2003-02-20			
; PRIOR APPLICATION NUMBER: 60/191,078			
; PRIOR FILING DATE: 2000-03-21			
; PRIOR APPLICATION NUMBER: 60/206,848			
; PRIOR FILING DATE: 2000-05-23			
; PRIOR APPLICATION NUMBER: 60/207,727			
; PRIOR FILING DATE: 2000-05-26			
; PRIOR APPLICATION NUMBER: 60/230,335			
; PRIOR FILING DATE: 2000-09-06			
; PRIOR APPLICATION NUMBER: 60/230,347			
; PRIOR FILING DATE: 2000-09-09			
; PRIOR APPLICATION NUMBER: 60/242,578			
; PRIOR FILING DATE: 2000-10-23			
; PRIOR APPLICATION NUMBER: 60/253,625			
; PRIOR FILING DATE: 2000-11-27			
; PRIOR APPLICATION NUMBER: 60/257,931			
; PRIOR FILING DATE: 2000-12-22			
; PRIOR APPLICATION NUMBER: 60/267,636			
; PRIOR FILING DATE: 2001-02-09			
; PRIOR APPLICATION NUMBER: 60/269,308			
; PRIOR FILING DATE: 2001-02-16			
; Remaining Prior Application data removed - See File Wrapper or PALM.			
; NUMBER OF SEQ ID NOS: 78614			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 70176			
; LENGTH: 2398			
; TYPE: PRT			
; ORGANISM: Staphylococcus aureus			
; FEATURE:			
; NAME/KEY: MISC FEATURE			
; LOCATION: (311)..(311)			
; OTHER INFORMATION: X-any amino acid			
; FEATURE:			
; NAME/KEY: MISC FEATURE			

LOCATION: (313)...(313)
OTHER INFORMATION: X=any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (327)..(327)
OTHER INFORMATION: X=any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (328)..(328)
OTHER INFORMATION: X=any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (336)..(336)
OTHER INFORMATION: X=any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (350)..(350)
OTHER INFORMATION: X=any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (362)..(362)
OTHER INFORMATION: X=any amino acid

US-10-282-122A-70176

Query Match 3.7%; Score 189; DB 15; Length 2398;
Best Local Similarity 18.7%; Pred. No. 0.03;
Matches 210; Conservative 153; Mismatches 444; Indels 314; Gaps 46;

QY 2 KHLKTVALTITVSVTHN-----QEVSILVKPEILKQTQAASSISIGADYAESGSKS 55
||| ||| : : : : :
1111 KOHLG---TLTHITTAQRNDLTNRISQATNLAVGESVKQN--ANSLDGA-----MGNLQ 1159

QY 56 LKINETSQPV-----DDVTDLFSDKRTTPEIKD-----NLAKPREOELKAV--- 99
||| ||| : : : : :
1160 TALNDSKGILASQNFILDQEQRNAYNAQVASAEIILNKQTPKTAKTAVEQALNNVNA 1219

QY 100 -----TENTESEKQ-----ITSGSQLEQ-SKESL--SLNKTFPSTSNWEICDFITKGT 145
||| ||| : : : : :
1220 KHALNGTONLNNAKAAITAINGASDLNQKQKDALKAQANGAQORVSNADVOHNATELNT 1279

QY 146 LVGLSKGV-----EKLSDTDHLV-----LPQOADGT 173
||| ||| : : : : :
1280 AMGLAHAIADTKNTLASSKYVNADSTKQNAVTTKVTAHEIISGTPVVVTPSEVATAA 1339

QY 174 QLIQVASFAPTPD-----KKTAIAEYTSRAGENGESQLDVD 210
||| ||| : : : : :
1340 NQVSAKSELNGDERUREAKQNAVDAIDALTQNPQRAKUKEQVGQANRLVEDVQTQTN 1399

QY 211 GREINEGEVFNSYLKKVTIPTGYKHGGQDAFVDNKN-----IAEVNLPESL 258
||| ||| : : : : :
1400 GOALLNAMKGLRDSIANETTVKTSQNY--TDASPNNQSYNSAVSNAGKIINOINNPN--- 1454

QY 259 ETISDVAFPAHLAKQIDLDPNLKAIGELAFFNOITGKLSL-----PROLMRLARAF 311
||| ||| : : : : :
1455 -TWDTSAITQATTQVNVNAKNGLAENLR--NAQNTAKQNLTLTSHLTNNOKSAISSQID 1511

QY 312 KSNHIKTIETPRGNSLVIGEASPDNDLSOLMLPGOLEKIBSEAPT--GNFPDHHYNRV 369
||| ||| : : : : :
1512 RAGHVSEVATKNAAT---ELNTQMGNEQAHDQNTVK-QSVKFTDADKAKRAYTNAV 1567

QY 370 -----VLWTKSGNQPSGLATENTYVPDKSLWOESPEDYTKWLFEEDFTYQKNSVTGSFN 424
||| ||| : : : : :
1568 SRAEAILTQGNANTSQODVEAIIQNVSSA-----KNALNGDON 1606

QY 425 KGLQKVRKNKULIPIKQHGVTITEIGDNFAFRVDFONKTRUKYDLEBKLPSTIRKIGA 484
||| ||| : : : : :
1607 -----VTNAKN-----AAKNALNLT SINNAOKRDULTTKIDQATTVAGVEA 1647

QY 485 FAFQSNNLKSFASDDLBEIKGAPANNRIETLELK-----DKLVTIGDAAFHINHYY 537
||| ||| : : : : :
1648 VNSTQTQMTAMAN-----LONG--INDKINTLASENTHDDSDCKKTYATQAVTNAENIL 1700

QY 538 -----AIVLPESVQBGRSAPRQNGANNLIIPMGSKVKTGLGEMAFNLSNRLBHLDSL 588
||| ||| : : : : :

```
Db 237 NLTEKLIATNEHEKDEKQIILNSHSHSUKENFISILETELKRVDRSLDQRLDVLST 296
Qy 310 AFSNHIKTIPTFGNSLKVIGEASFODNDLSQMLPDGLEKTESEAFPTGPGDDHNNRV 369
Db 297 KOKENQALLEYKSTHK-----QEDSIK--TLEKGLTILSQK---KXEDGIN--- 341
Qy 370 VLWTKSGKPNPGLATENTYVNPDP--KSLWQESPIDY-----TKWLEEDFTYQKNSVTGFS 423
Db 342 ----KMGKOLFALSREWQAVEENCKNLQKEDKSNVNHQKETSLSKEDIAAKITEIKAI- 396
Qy 424 NKGLQKVK--RNKULEIPKOHNGVTIITEIGD-----NAFRNVDFQW 462
Db 397 NENLEMKIQCNLKEKHEKISKELVYKSRFOSHNDNLVAKTEKLKSLANNYKQMAEN 456
Qy 463 KTLRKVDLEKVLKPSRIR-----KIGAFAPQSNLKL-----SPEAS-----DLEBI 504
Db 457 ESILIK-AVESKNESIQIOLNQNKKDSMSQENFQIERGSEKIEKNIQKKTISLEQOT 515
Qy 505 KEGAFMNNRIETLE-----LKDKLVTIGDA-AFHINHIYAIVLPEVQIEGRSAFRON 556
Db 516 KEBIISKSDSSKDEYESQISLLKEKLETATTANDENVNKISELTKTREBLEAELAAAYK-N 574
Qy 557 GANNL-IFMGSKVKTIGEMAFSLNRLEHLDLSEKQKLTPLVQAFSDNALKEVL----- 609
Db 575 LKNELETKLETSEKALKVEKNEHEHLKEEIKIQLEKEATETKQOLNSLRANLESLEKEHEB 634
Qy 610 LPASLKTIREEAPKNHLKQLEVASALSALSHAFNALDNDGDQFDMKVVVKTHNSYALA 669
Db 635 LAAQLKYEQIANKERQYNEEISQ-----LNDEITSTQOENESIKKKNDE-----L 681
Qy 670 DGHFTIV-----DPDKLSSTIVDLEK-----ILKLEIGLDYSTL----- 703
Db 682 EGRVKAMKTSBQSNLKSSEIDALNLQIKELKKNETNEASLESIKSVESETVKIKEL 741
Qy 704 -----RQTQTQFRMTAGKALLSKSNLRQEKOKFLQ-----EAOFFLGRV-DLD 749
Db 742 QDECNFKKEVSELDKASEDKNSKYLELOKESEKIEELDAKTTELKIQLEKITNLS 801
Qy 750 KAKAEKALVTYKAT-----KNG-QLLERSINKAVLAYNNSAIKAVNKRLEKELDLIT 803
Db 802 KAKESESELSRLKTSSEERKNAEBOLEKLNKNE--IQIKNQK-----FEKERKLIN 851
Qy 804 GLVEGKPLAQATWQGVYLLTKPLPEYIIGLVVYFDKSGKL-----IYALDMSDTIG 858
Db 852 ---EGSSTIQ-----PYSEKINTLEBELRLQENELKAKEIDNTRS 891
Qy 859 EGOQDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKTILNSKLSQTSI-----ROVP 913
Db 892 ELEKVSLSNDELLEEKQK-----TIKSLOD-EILSYKDKITRDNDEKLLSIERNKRDLE 944
Qy 914 TAAYHRAGIFOALQNAEAEQALLPKPGTHSEKSSSSESA---NSKD--RGLQSNPKTN 967
Db 945 SLKEQ----LRAAQAKSAKVEBLKK-----LEBESSKAELEKSKMMKLESTIESN 995

RESULT 70
US-10-282-122A-60608
; Sequence 60608, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60608
; LENGTH: 2013
; TYPE: PRF
; ORGANISM: listeria monocytogenes
US-10-282-122A-60608

Query Match 3.7%; Score 188.5; DB 15; Length 2013;
Best Local Similarity 21.9%; Pred. No. 0.025;
Matches 250; Conservative 148; Mismatches 429; Indels 315; Gaps 66;

Qy 18 VTHNEVPSLVKEPILKOTQASSISGADYABSSGSKLKNETSGPVDOT--VTDL--- 72
Db 975 ITLNEAIVPTDAPRDEIAYNSIAYRIDKETGSGTSLASEPPRGVKTNPSTDLNIA 1034
Qy 73 ---FSDKRTTPBKIKDNLAKEPREQLKAV-----TENTESEK---QITSGSQLEQSK 120
Db 1035 GNSFSD-----LNKNGVKDADLGLNAVKLDLYKKNNEPEKVEVTYSSDALDK 1087
Qy 121 SLSLN-----KTVPSTSNWEICDFITKG-NTLVGLSKSG-----VEKLSQTDHLVLP 168
Db 1098 LPDFNGLNNGTYKIAAHLPN-KNADFITTPGNKIVKDSKDDSIGHITVNNSTFTI--DD 1144
Qy 169 AADGTQL---IQVASFAFTP-----DKKTAIAEYTSRAGENGESIQLDVDGKEI 214
Db 1145 LANGNPKNLVGLQIPIYATPIKGAVFVNKDGEPKITSY-----GQDYHIALFDKDG 1200
Qy 215 -----INEGEV-FNSYLLK-----KVTIPTGYKHIGQDAFVNDKNIAEYNLPES 260
Db 1201 QSAIKTNSKGPSFNDVAIKNPADFKLVTAFTGTFV-----YSAKN----- 1243
Qy 261 ISDYAPAHALAKQIDLPDNLKAIAGELAFDQITKLSLPLQMLRAERAFKSNHIKTI 320
Db 1244 ----PLFNMSTKEYTLNSVPGVGGA--EYIIT-ETSKPTTKI-ILDKAVTPNAI-TIE 1294
Qy 321 PRGNSLKVIGEASFODNDLSQMLPDGLEKTESEAFPTG-----PGDDHNNRVVLWTK 376
Db 1295 SSDEATEVTNWTLESDSGTVVY-----TGTGNTIRIPNDE-----GTIYA 1335
Qy 377 KNPSGLATENTYVNPDKSLWQESPIDY---TKWLEEDFTYQKNSVTGFSNKGLOKVRN 433
Db 1336 KNTATDEAGNT-ASDEKTF-----DIDYTVPLTVNQDASAEVNSTE--ANIGWIK--P 1384
Qy 434 KNLKIPKOHNG-----VTITEIGDNA-----FRNVDFQNTK- 464
Db 1385 LNVAATDTHDGNITPVDVYSKVKWDVLGTPVTVTATDASGNKATQTVNLRIVDTTSP 1444
Qy 465 -----LRKYDLEEVKLPSTIRKIGA-----FAFQSNLKS 494
```

```
Db 1445 LITNPLTYSIENMRK-LTEQELTYAAGLIGDNDYDLAPGAQVQPNKQPMVFTSNFSTI 1502
Qy 495 FRASDDLEIEKGAFM-----NNRIE---TLEKDKLVLTIGDAAFHINHIIYAIVL 541
Db 1503 F---GDIAVKAQGVQVQVNLADSSGQAIPQITITINVVD---TVGPVKAQDNVSYHVT 1556
Qy 542 PESVOEIGRSAPRQNGANNLIIPWGSKVTLGEMAFSLNLEHLDLSE--QKQLTPIPVQA 599
Db 1557 TKTEAEFFQDARLDVTDNN-----DDTDLIITSNPAEKVNLNPKGVEVITATDT 1608
Qy 600 PSDNALKEVLLPASL-KTIREBAFKKNHLKQLEVASA--LSHIAFNAFLDDNDGDFQFNK 656
Db 1609 KGNQITKEITVQVSKDKPVITADPKISVQKIEVTEANFLSGVHAEVTDDELDCDVKITSD 1668
Qy 657 VVVVTHN---SYALA---DGEHFIVDPDKLSSITVDLEKILKILIEGLDYSTLRQTQT 709
Db 1669 PAEKVDNFNKGVTYVTLNAKDEYGNATBEPKVSYSI--FNKIAPTNNADNKTIEAVNEL 1726
Qy 710 QPRD--MTTAGKALLSKSNLRQGEKQFLOBAQFPLGRVLDLDAIAKAEKALVTTKATK- 766
Db 1727 PLSIESIFKIEADYLSGNKL-----KVYTPPE-QTIKGNVPGEYSI-----KVTTKD 1772
Qy 767 -NGQLLERSINKAVLAYNNSAIKKA-NVKRLB---KELDLL-----TGLVEG---KGP 811
Db 1773 DSGNTAEVTVITIKDTGPSIMTKSTYKLEVOQSEPNWITFFGIGKATDIVDGDVTKNI 1832
Qy 812 LAQAATWQV---GVYLLKTPPLPYIYIGLVNYYFDKSGKLIYALDMSDTIGSGQDAYGN 867
Db 1833 KVDSEVNLNKVGTY-----PIVFTVDALGNESKLSKSTVQIVDT-----SS 1875
Qy 868 PILVVDENEGHALAVATLADYEGI-DIKT-ILNS--KLSQTSIRQVPTAAHYHRAGIF 923
Db 1876 PELTIDKLEISYP--TGKITSQFLQDQIGTKVTNSYGTGVKVTNNLSKI--VDMNKAGKY 1931
Qy 924 QAIONA-----AABAEOLLPPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAIILPR 977
Db 1932 KVTVTATNSGGVGAETILL-----TVKNSDSSFIAVPSKD-----DNKNKPAKNIPK 1979
Qy 978 TG 979
Db 1980 TG 1981
```

RESULT 71

US-09-815-242-5816
Sequence 5816, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Cart, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-05-23

PRIOR FILING DATE: 2000-05-23

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-10-23

PRIOR FILING DATE: 2000-11-27

PRIOR FILING DATE: 2000-11-27

PRIOR FILING DATE: 2000-11-27

PRIOR FILING DATE: 2000-11-27

PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5816
LENGTH: 2478
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match 3.7%; Score 187.5; DB 9; Length 2478;
Best Local Similarity 19.3%; Pred. No. 0.039;
Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

Qy 12 LTTVVVTHNQEVFSLVKEPILKQTOASSSSISGA-----DYABSSGSKLKINETSGPV 65
Db 1170 ITABEIAAANADVNAV-----TQANSNIEAANSQNDVDQAKTTGENSI----- 1213

Qy 66 DDTVTDLPFSDIKRTPPEKIKONLAGPREQELKAVNTESKQITSSQLESQESLSLN 125
Db 1214 -DQVTPTVNKKATARNETAIL--NNKLEIQATPDATDEEKQ---AADAENATENGKAN 1267

Qy 126 KTVPTSTNWEICDFTTKGNTLVGLSKSGVVKLSQTDHLVLPQADGTLQIVASFATP 185
Db 1268 QAIISAAT-----NAQVDEAKANAE-----AAINAVTP 1295

Qy 186 D--KRTAIAEYTSRAGENGEISQLDVDGKEIINEGEVFNLSYLLKKVVTIPTGYKHIGQDAF 243
Db 1296 KVVKKQAAD-----EIDQLQATQTVINN-----DQNAI 1325

Qy 244 VDNKNAEVLNPELETISDYAFPHLAKQIDLDPNLKAIGELAFPNQITGKLSLPRQL 303
Db 1326 TEKK-----EAAIQOLATAVTDKANNITA-----ATDDNGV----- 1356

Qy 304 MRLAERAFKSNHIKTIERFGNSLVIGESAFQDNDLSQLMLPDGLEKIESEAFNGNPGDD 363
Db 1357 -----DQAKDAGKNSIQSTQPATAVKSNKANDVDQAVTTQNOAID 1396

Qy 364 HYNRRVVLWTKSGKGNPSGLATENTYVNPDKSLMOWESPEDYTKWLEEDFTYQKNSVTGFS 423
Db 1397 -----NTTGATTEK--NAAKDLVLKAKEKAYQDILNAQT---NDVTQIK 1437

Qy 424 NKGLQKVRKNKLEIPKHNGV-----TITEIGDNAPRNVDPONKTL--RKYDLSEVKLPS 477
Db 1438 DQAVADIQ-----GITADTTIKVDKDELATKANEQKALIAQTADATTEEKEQ 1485

Qy 478 TIRKIGAFAPQSN-NLKSFRASDDLBEIKEGAFM-----NNRTETL-ELKDK 522
Db 1486 ANQVVDLQLTQGNQNIENASQIDVDNTAKONAIQAIIDPIQASTVDVKNARAELETQMNK 1545

Qy 523 LVTIGDAAFHINHIYAIVLPESVOEIG--RSAPRQNGANNLIIPMGSKVYKTLIGEMAFLSNR 580
Db 1546 ITEILNNNETTNE-----EKGNDIGPVRAAY--EGLNNI-----NAATTG----- 1585

Qy 581 LEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREBAF-KKNHLKQL----- 630
Db 1586 ---DVTTAKDTAVQKVQQLHANPVKK---PAGKKELDQAAADKKTQTEQTPNASQOEIN 1638

Qy 631 ---EVASALSHIAFNALDDNDGDEQPDNKV-----VVKT--HNSYALADGEHF-- 674
Db 1639 DAKQEVDTLNOAKTN--VDQSTWNEYVDNAVKEGAKINAVKTFSEYKDKALAKIEDAYN 1697

Qy 675 --IVDDP-----KLSSTIVDL--EKILKLEGLDYSTLRQTQ-----TQFRDMT- 715
Db 1698 AKVNEADNSNASTSSEIAEAKQKLAELKQTDQNVNQATSKDDIEVOIHNDLDINDYTI 1757

Qy 716 TAGKA-----LLSKSNLRQGEKQKFL-----QEAQFELGRV-----D 747
Db 1758 PTGKESATTDLYAYADQKKNISADTNAQODEQQAQKQVDQNVQTALESINNGVDNGD 1817

Qy 748 LDKAIAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKKANVVKRLEKELDLITG--- 804

Db 1818 VDBALTQKAAI---DAIQVDAATVKKANQAI-----EVKAEDTKESIDQSQDLTAEEK 1868
QY 805 -----LVBEKGKPLAQATMVQGVYLLKTLPLPEYIIGL-----NVVFKSGK---- 846
Db 1869 TEALAMIKQITDOAQKQITDATTAEVEKAKAQ-----GLEAFDNIQIDSTEKQKAI 1920
QY 847 --LIYALDM-----SDTIGEGQKDAYGNPILNVDENEGYPHALAVATLADY--EGLDI 895
Db 1921 BELETALDQIEAGVNVNADATTE--EKEAFTNAL-----EDILSKATEDISDQTTNABE 1972
QY 896 KTLNSKLSOLTSIROVPTAAVHRAGI-----FQAIONAAAEAE----- 934
Db 1973 ATVKNBALEQLKQRIINPEVKNKALEAIREVVNKQIEIINKADADASAKAEIARTDLGRYF 2032
QY 935 -----QLLPKPGTHSEKS-----SSSESANSKDRGLOSQPKTNRGRHSAIL 975
Db 2033 DRFADKLDKTQTNAAEVLQNVTPAIEAIVPQNDPDANDTNGINDNDATANSANATP 2092
QY 976 PRTG 979
Db 2093 ENTG 2096

RESULT 72

US-09-815-242-12967
; Sequence 12967, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12967
; LENGTH: 2478
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12967

Query Match 3.7%; Score 187.5; DB 9; Length 2478;

Best Local Similarity 19.3%; Pred. No. 0.039;
Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

QY 12 LTTVSVVTVNQEVSILVKEPILKQTOASSISGA-----DYAESGSKLAKINETSGPV 65
Db 1170 ITAEETAANADVNAV-----TQNSNIEANSQNDVDQAKITGENSI----- 1213
QY 66 DDTVTDLDFSDKRTTPKIKDNLAKPQEQLKAVTENTSEKQITSGSQLEQSKESLSLN 125

Db 1214 -DQVTPTVNNKATARNEITAIL--NNKLOEIQATPDATDEEQ---AADAEBANTENGKAN 1267
QY 126 KTVPTSTNWEICDFITKGNLTGLSKSGVSKLSQTDHLVLPQAADGTQLIQVASFATP 185
Db 1268 QAIASAATT-----NAQVDEKANA-----AANAVATP 1295
QY 186 D--KKTATABYTSRAGEGEISQLDVDGKBIINEGEVFNYSYLLKKVTIPTGYKHIGQDAF 243
Db 1296 KVVKQQAAD-----EIDQLQATQTNVNN-----DQDAT 1325
QY 244 VDNKIAEVLNPLESETISDYFAFAHLALQIDLPNLKAIAGELAFDNOITGKLSLPROL 303
Db 1326 TEER-----EAAIQQLATAVATKKNITA-----ATDDNGV----- 1356
QY 304 MELAERAPKSNHIKTIEPRGNSLKVIGEASFQNDLSQMLPDGLEKIESEAFNGPGDD 363
Db 1357 -----DOAKGAGKNSIOSTQPATAVKNAKNDVDQAVTTQNOAID 1396
QY 364 HYNRRVWLWTKSGKNPSGLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFS 423
Db 1397 -----NTTGATTEK--NAAKOLVLVAKAKAYQDILNAQTT---NDVTQIK 1437
QY 424 NKGLOKVKRNKONLEIPKQHNGV-----TTEIGDNAPRNVDPONKTL--RKYDLEEVKLS 477
Db 1438 DOAVADIQ-----GITADTTIKDVADELATKANQKALIAQTADATTEKEQ 1485
QY 478 TIRKIGAFAFQSN-NLKSFEASDDLEEIKEGAFM-----NNRIETL--ELKDK 522
Db 1486 ANQQVDAQLTOGNQNIENQAQSIDDVNTAKDNAIQAIQIDPTQASTDVKTNARABELTEMQNK 1545
QY 523 LVTIGDAAAFHINHIYAIVLPEVQBIG--RSAFRONGANNLIFMGSKVKVTLGEMAFLSNR 580
Db 1546 ITEILLNNNETNE-----EKGNDIGPVRAAY--EGLNNI---NAAATTTG----- 1585
QY 581 LEHLDSQKQLTEIPVQAFSDNALKEVLLPASLTIREAF--KQNLKQL----- 630
Db 1586 ----DVTTAKDTAVQVQOLHANPVKK---PAGKELDQAAADKKTQIEQTPNASQOEIN 1638
QY 631 ----EVASALSHIAFNALDDNDGDGFQNKV-----VVKT--HNSYALADGEHF-- 674
Db 1639 DAKQEVDTLNOAKTN--VDQSTNEYVDNAVKEGKAKINAVKTFSEYKDKALAKIEDAYN 1697
QY 675 --IVDPD---KLSSTIVDL--EKILKLEGLDYSLRTQTT-----TQFRDMT-- 715
Db 1698 AKVNEADNSNASTSSEIAEAKQKLAELKQATADQNVNQATSKDDIEVQIHNDLDNINDYTI 1757
QY 716 TAGKA-----LLSKSNLRQGEKQKFL---QEAQFFLGRV-----D 747
Db 1758 PTGKESATTDLYAYADQKKNISADTNTATQDEKQQAIKQVDQNVQTALESINNGVDNGD 1817
QY 748 LDKATAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKKANVKRLEKELDLLTG--- 804
Db 1818 VDDALTOGKAAI---DAIQVDAATVKKANQAI-----EVKAEDTKESIDQSQDLTAEEK 1868
QY 805 -----LVBEKGKPLAQATMVQGVYLLKTLPLPEYIIGL-----NVVFKSGK---- 846
Db 1869 TEALAMIKQITDOAQKQITDATTAEVEKAKAQ-----GLEAFDNIQIDSTEKQKAI 1920
QY 847 --LIYALDM-----SDTIGEGQKDAYGNPILNVDENEGYPHALAVATLADY--EGLDI 895
Db 1921 BELETALDQIEAGVNVNADATTE--EKEAFTNAL-----EDILSKATEDISDQTTNABE 1972
QY 896 KTLNSKLSOLTSIROVPTAAVHRAGI-----FQAIONAAAEAE----- 934
Db 1973 ATVKNBALEQLKQRIINPEVKNKALEAIREVVNKQIEIINKADADASAKAEIARTDLGRYF 2032
QY 935 -----QLLPKPGTHSEKS-----SSSESANSKDRGLOSQPKTNRGRHSAIL 975
Db 2033 DRFADKLDKTQTNAAEVLQNVTPAIEAIVPQNDPDANDTNGINDNDATANSANATP 2092
QY 976 PRTG 979
Db 2093 ENTG 2096

```
RESULT 73
US-10-470-048B-220
; Sequence 220, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; FILE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
; CURRENT APPLICATION NUMBER: US/10/470,048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 220
; LENGTH: 2478
; TYPE: PRN
; ORGANISM: Staphylococcus aureus
US-10-470-048B-220

Query Match      3.7%; Score 187.5; DB 17; Length 2478;
Best Local Similarity 19.3%; Pred. No. 0.039;
Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

QY      12 LTTVSVTHNQSVFLVKEPILKQTOASSISGA-----DYAESSGSKLKNINETSQPV 65
DB      1170 ITABEIAANADVDAV-----TQANSNIEAANSQNDVDQAKTTGENSI----- 1213

QY      66 DTVTDLFSDKRTTPEKIDNLAQPREQELKAVTENTSEKQITSGSQLEOKESLSLN 125
DB      1214 -DQVPTVNNKATARNETAIL--NNKLEIQATPDATDEKQ---AADAENTENGKAN 1267

QY      126 KTVPTSNWEICDFITKGNLTVLGSKGVKLSQTDHLVLPQAAQDTGLIQVAFATP 185
DB      1268 QAISAATT-----NAQVDEAKANAE-----AANAVTP 1295

QY      186 D--KKTATAYTSRAGENGHEISQLDVKGKEIINEGEVFNYSLLKKVTIPTGYKHIGQAF 243
DB      1296 KVVKQAAKD-----EIDQLQATQTNVNN-----DONAT 1325

QY      244 VDNKNIAEVLNPLESLETISDYAFALHALKQIDLPNLKAIGELAFPDNQITGKLSLPRQL 303
DB      1326 TEEL-----EAAIQLATVTDKKNITA-----ATDDNGV----- 1356

QY      304 MRLAERAFKSNHIKTIEPRGNSLVIGESFQDNLSQLMLPDGLEKIESEAFPTGNPGDD 363
DB      1357 -----DQAKDAGKNSIQSTQPATAVKSNKXNDVDQAVTTQQAID 1396

QY      364 HYNRRVVLTKSGKPSGLATENTVYNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGPS 423
DB      1397 -----NTTGATTEK--NAAKOLVLKAKKAYQDILNAQTT---NDVTQIK 1437

QY      424 NKGLOKVKRKNLEIPKQHNGV---TITEIGNAPRNVDFQNKTL--RKYDLEEVKLPS 477
DB      1438 DOAVADIO-----GITADTTIKDVAKDELATKANEQKALIAQTADATTEKEQ 1485

QY      478 TIRKIGAFAPQSN-NLKSFEASDDLEIEKEGAPM-----NNRIETL-ELKDK 522
DB      1486 ANQVDAQLTQGNQNIENASIDVNTAKDNAIQDIPQASTDVKTNARAEELLTEMQNK 1545

QY      523 LVITGDAAPHIHHIYAIVLPESVQIG--RSAFRONGANLLFPMGSKVKTLGEMAFLSNR 580
DB      1546 ITEILNNNTTNE-----ERKNDIGPVRAAY-EEGLNNI-----NAATTTG----- 1585

QY      581 LEHLDLSEKQLTEIPVQAFPSNALKEVLLPASLTKTIREAF-KKNHLKQL----- 630
DB      1586 -----DVTAKDTAVKQVQLHANPVK---PAGKELDQAAADKKTQIEQTNASQOEIN 1638

QY      631 -----EVASALSHAFNALDNDGDEQDNKV-----VVKT--HHNSYALADGEHF-- 674
DB      1639 DAKQEVDTLQNAKTN-VQDSSTNEYVDNAVKEGRKAKINAVKTFSEYKKDALAKIEDAYN 1697
```

```
QY      675 --IVDPD--KLSSTIVDL-EKILKLEGLDYSLRQTTQ-----TQPRDMT- 715
DB      1698 AKVNEADNSNASTSETAEAKQKLAELKQTADQNVNQATSKDDIEVQIHNDLDNINDYTI 1757

QY      716 TAGKA-----LLSKSNLRQEKQKFL-----QEAQPFPLGRV-----D 747
DB      1758 PTGKESATTDLYAYADQKKNISADTNATQDEKQALKQVDQNVQTALESINNNGVDNGD 1817

QY      748 LDKAIAKAKALVTKATNGQOLLERSINKAVLAYNNSAIKKANVKRLEKELDLITG--- 804
DB      1818 VDDALTQGKAAI---DAIQDQATVKPKANQAI-----EVAEDTKESIDQSDQLTABEK 1868

QY      805 -----LVEGKGPLAQATMVQGVLLKTPPLPEYYIGL---NVYFDKSGK--- 846
DB      1869 TEALAMIKQITDQAKQGITDATTTAEVEKAKAQ-----GLSAFNIQIDSTEKQKAI 1920

QY      847 --LIVALDM-----SDTIGEGQKDAYGNPILNVDEDNEGYHALAVATLADY-EGLDI 895
DB      1921 EELETALDQIBAGVNVNADATTE-EKEAPTNAL-----EDILSKATEDISDQTTNAEI 1972

QY      896 KTIINSLKSLQTSITRQVPTAAHYRAGI-----FOAIQNAABAE----- 934
DB      1973 ATVKNSALEQLKQAKRINPEVKKNALAEIREVNVNQIIEIKNADADASAKETARTDLGRYF 2032

QY      935 ----QLLPKPGTHSEKS-----SSSESANSKDRGLQSNPKTNRGRHSAIL 975
DB      2033 DRFADKLDKDTOTNAEVAELQNVTPITAEIIVPQNDPDANDTNGINDNDATANSNANATP 2092

QY      976 PRTG 979
DB      2093 ENTG 2096
```

```
RESULT 74
US-10-080-334-234
; Sequence 234, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkets, Richard A
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kerkuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zehusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
```

PRIOR APPLICATION NUMBER: 60/278,796
PRIOR FILING DATE: 2001-03-26
PRIOR APPLICATION NUMBER: 60/281,521
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/276,677
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/311,595
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/270,220
PRIOR FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: 60/274,295
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/318,526
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/286,548
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/291,765
PRIOR FILING DATE: 2001-05-17
PRIOR APPLICATION NUMBER: 60/270,797
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: 60/276,400
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/270,810
PRIOR FILING DATE: 2001-02-23
NUMBER OF SEQ ID NOS: 388
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 234
LENGTH: 1348
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-10-080-334-234

Query Match 3.7%; Score 187; DB 15; Length 1348;
Best Local Similarity 18.6%; Pred. No. 0.018;
Matches 219; Conservative 159; Mismatches 344; Indels 454; Gaps 55;

QY 92 RQELKAVTENTSEKQITSGSQESLSLN-KTVPSTSNWEICDITTKGNTLVGLS 150
DB 11 RQRELAQRSPKQKTAKLPSSQATSKCNLQAAPKTT-ENTF-GQL 60
QY 151 KSGVEKLSQTHLV-LPSQAADGTQIQVASFAPTPDKKTAIAEYTSRAGEN-201
DB 61 KLTIBELDSLNNLRIRRIPEKAFDGLK-DSLNELRANLLGDNLPISPT 109
QY 202 -----GEISQLDVQKEI--INEGEVFNYSLLKKVTIPTGYKHIGQDAFVNKIAEY- 252
DB 110 AELHVLKNRLDLSGNKIKLIEG-----LLK-----GCMDL-KEYIIDRNSLTSP 156
QY 253 ---NLPESETISDYAFALHAKQIDLPDLNKAIGELAFFNQITGKLSLPRQLMLAE 308
DB 157 TNSLNGPSALR-----HLSLRO-----NQI-GSL-----LAD 182
QY 309 RAFKSNHIKTIIFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFTEGFGDDHYNNR 368
DB 183 SFNAQRQLEIIDLNRHNVIRISDLSAPK-----GLQKIREIKLAG-----NR 223
QY 369 VVLWTKSGKNPSGLATENTYVNDK-SLWQESPEDYTKWLEEDTYQKNSTGFSNKG 427
DB 224 I-----SHLNSDVEKLSQSLQKLD-----LSENFPGQPTVALAAVPL 262
QY 428 QKVKRNKL--BIPKQNGVTITEIGDNFRNVDFQNKTLRYDLEEVKLPSTIRKIGAF 485
DB 263 KHNLSSNMLOQDTHYMQV-----RSLESDLSRNTITI-----TPGTFRMGAL 310
QY 486 AFQSNLSKFEASDDLEIEKEGAFMNRIETTELKDK--LVTIGDAAFHINHIYVILPE 543
DB 311 KYLDLSLNSLRITIED--DALEGL--DSLQTLIIKDNILLVPGSALGRLPOLTSIQLDY 365
QY 544 S-----VOETGRSAFRO-NGANNILPFGSKVKTIGEMA 575
DB 366 NRVAALSABILGSLQAGDITTLSSRNVIRELPFGSFQMFSSLHTLDLSGNSLAVINADT 425
QY 576 P--LSNRLEHLDLSEK-----QLTEIPQAFSDNALKEVLLPAS-----LKTIRE 620

DB 426 FAGLESTLWALKLSQNRLTGLGGAPWLPBLSLDSGNTLTE--LPSTIFIELENVQSL 483
QY 621 AFKQVHLKQL-----EVASALSHIAFNALDNDGDQPD 654
DB 484 NLSGNHLTFLTGALFKPLDLQVIDLSCNIRQISGDLIAGLQDLKHIYINDNQLEQD 543
QY 655 NKVW-----VKTHNSYALADGHEF-----674
DB 544 GSFVNLWNISSIDLNNRIGSIRSAFVNVWMLQKLDLHGNQLSAFKGYFTGTGIBEL 603
QY 675 -IVD-----PDKLSSTIVDLEKI-----LKRIEG 697
DB 604 DISDNQLSVLPSPSPRIHPRLREIRAANNKFSFFPAELISTIQLEHIDLSHNLKTIEE 663
QY 698 LDYSTL-----RQTTQTQFRDM-----TTAGKALLSKS 725
DB 664 LDFARLPRLRVLLVANNQLDVMVSEMAFNSTQQLDLAHNNLDRIGERTFGLVRLEQL 723
QY 726 NLRQG-----EKQKF-----LQEAQFFELGRVLDL--KA 751
DB 724 NLEGNRLSELSDGVFERTKLOMLENINLAHNRFEYAPLNALQRFVSSVDLSHNKIKE 783
QY 752 IAKAEKALVTKATKNGQLLERSINKAVLAYNNSAIKKA-----NVKRLKELDLTLGLVE 807
DB 784 LPGDDSIW-----NIKRIDLSPNPLSSKAVHNVLEPKTVRELSL-----824
QY 808 KKGPLAQATMVQGVYLLTPTPLPYIYIGLVNVPKSGK-----LIIVALDMSDTI 857
DB 825 -----ACTGIENLELETP-----FLOFLNLSHNKLKXNVKPEVFORVTLTDLSSNQ 873
QY 858 GEGOKD-AYGNPILNV-----DEDNEGYHALAVATLADYEGLDIKTILNSKLSQLSI--- 909
DB 874 LESUEDLSNAMPQLQVLOSLDVSNNFSFVSGNFKLEML--RSLRLSHLPQCTRIEKN 931
QY 910 --ROVPTAAYHRA-----GIFQ-----AIONAAAEABQLLP-----938
DB 932 AFKQPLNVLSEAYDLPLGVLQDLSQILELPLGVLVIEVKDSSIGSEQIQPLKHPRLK 991
QY 939 KPGTHSEKSSSESANSKDRGLQSNPKTNRGHSAI 974
DB 992 SLGIRGDRLKISSTGTL--AGLKSNDLSVQLRNTSL 1025

RESULT 75
US-10-369-493-1950
; Sequence 1950, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xiaofeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1950
; LENGTH: 1312
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1950

Query Match 3.7%; Score 186.5; DB 15; Length 1312;
Best Local Similarity 19.0%; Pred. No. 0.018;
Matches 200; Conservative 174; Mismatches 362; Indels 315; Gaps 49;

QY 21 NOEVPSLVKEPILKOTQASSISGADYAESGSKLKINETSGPVDVDTVDLFSDKRTPP 80

Db 800 -TOSTSKESLSNELTELKIAAKK---EQACKGEEDNLARLKKELTETELALKEAKEDLS 855
QY 638 HIAFNALDNDGDQDFNDKVVVTHNSVALADGEHFI VDPKLSSTI--VDLEKILKLE 696
Db 856 FLTSEMSSTSGEELKEA-----AKHNDKTKTIELIALRRDQRIKLOH 901
QY 697 GLDYSTLRTOQTQFRDMTTAGKALLSKSNLRQGEKQKFLQBAQFFLGR--VDLOKAI-- 752
Db 902 GLD-----TYERELKEMRLYK-----QKTLKDBEVKGRMEVELDNLLQY 944
QY 753 -----AKAEKALVT--KKATNGOLLERSINKAVLAYNSAIAKANKVLEKEL 799
Db 945 LREYVLSPEGAKEQYLETDPPEARVRVKKLJAIEEL-----GTVNLGSIDEFERVN 998
QY 800 DLLTLGLVEGKGLAOA--FMVQGVLLKTPLP--LPEYVIGLVNVPDKSGKLIYA----- 850
Db 999 ERYKFLSEKEDLEAKNTLFQVIEEMDEMYKRFNDTFVQIRSHFDQVFRSLFGGRAB 1058
QY 851 ----LMSDTIGEGOK---DAYGNPILNVDNDEGHYHALAVATLADYEGLDIKTILNSKLS 904
Db 1059 LRLTPNDLLHSGVEIIAQPCKLQNLNLLSGGERALTAL-----LF 1103
QY 905 QLTSTROVP-----TAAYHRAGIFQATQ 927
Db 1104 SILKVRPVPFCVLDEVEAALDEANVFRFAQ 1133

RESULT 77

US-10-282-122A-46664
; Sequence 46664, Application US/10282122A
; Publication NO. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46664
; LENGTH: 1788

; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-282-122A-46664
Query Match 3.7%; Score 185.5; DB 15; Length 1788;
Best Local Similarity 19.2%; Pred. No. 0.033;
Matches 223; Conservative 158; Mismatches 397; Indels 383; Gaps 57;
QY 2 KKHAKTVALTTTSSVTHNQEVSLVKEPILKQTPQASSISGADYASSSGSKLKINET 61
Db 582 QRNKTTDALLKDIPTDTHELDVSIV-----VNASYDD-----KRLVTGT 625
QY 62 SGPVDVTVT-----DL-FSD-----KRTTPEKIDNKLAKGPREQELKAVTENT 103
Db 636 VN--NYTVTNKNGFDLQFNEDINSAYVITYTKPTNNVIKDGKIK-----NTVTADN 676
QY 104 ESEKQITSGSQLEQSGESLSLNT--VFSTSNWEICDFITKNTLVGLSKSV----- 154
Db 677 GSSKENEAGFOOQNIKSNNAETNYKDKTTTWT-----ITVNNNNYPLNNAIITDFDHG 732
QY 155 -----EKL-----SOTDHLVLPQAAADGTQLIQVASFAFPPDKKTAIAEYTS-- 196
Db 733 GLQLKDKKLEIKDGYDTLQAEITYVLDVTDKGFKTLTGTVQSNMT-----KTLIVKTTDF 789
QY 197 -----RAGE-----NGEISQLDVGKEIINEGEV--FNSYLLKKVVTPTGYKHIGQDAFVD 245
Db 790 DYTLESKTSFKNTGNLSWIDAGSNPOSQNVKVEADFPDPTTKA---NGYKGSYNAQTK 846
QY 246 -----NKNAEVLNPLESETISD-----YAFHALLKQIDLDPNLKAIGELAPF 289
Db 847 BITWIIIGFNNVNEIKDPYVIDVIQDKQKLVPGSIEVRHMLN--GSPDNARP-GD--- 899
QY 290 DNQITGKLSLPLQLMLAERAFKSNHKTISFRG----- 323
Db 900 -----AVPIEQYELEPTDKNKTLOVHFQKQINSPIYIIFKTSLDGELIQNTYKNE 951
QY 324 -----NSLK-----VIGEASFQDNDS--QLMLPDGLKEIESEAFNGP 361
Db 952 AELKDGSKI VNTLKGDAQVNGKGSFVTKKAVQDDNYINNSAINESQSTIADAVVTDPT 1011
QY 362 D-----DHYNNRVVLWTYSGKNPSGLATE--NTYVNPDKSLMQESPEIDY 404
Db 1012 DNQVLVEDSFHLYPTTVDYGN-----VTKDTANELKQGTDYKLTITTDNNTGKHFEIAP 1067
QY 405 TKWLEEDFTYQKNSVTG-----FSNK-----GLQKVKENK----- 434
Db 1068 LKKIDRAVILYRSLINADDKKSVNKAIAAGNQTVTKTETVETIEVRMSSGSGGSAT 1127
QY 435 -----NLEIPKQHN-----GVTTIT-----EIGDNAFRNV-----DFQNKTLRKDYLE 471
Db 1128 NGRGNLEIIVKVDNDKVKYPLSGAEFTLYDRGTGKTVIRKITTDDKQIAKFNLLKRDYLLK 1187
QY 472 EVKLPSSTIRKIGAFAPQSNNLKSPASDDLEIEGAFMNNRIETLELKDVLVTGDAAF 531
Db 1188 ETAP-----EGYVISWDLKQKIVELGSOBTTTYKLANK----- 1222
QY 532 HINIYAI VLPESVQEIGRSAFRQNGANNLIFMGSKVKTGLGEMAFSLNRLHLDLSEKQ 591
Db 1223 --KFVGKVLTKS--DDLKNKNTLQAVFTLDDKKKI-----ISEHEKLTNDQOQ 1270
QY 592 LTEIPVQAFSDNALKEVLLPASL---KTIREAPFKNHLKQLEVASALSHIAFNALDND 648
Db 1271 IT-----VDN-----LKPGTYVLOETTAPEHYKLS-----TPQFTIKED--- 1306
QY 649 GDEQFDNKVVKVTHNSYALADGBHFIVDPDKLSSTIVDLEKILK-----LIEGLDYSTL 703
Db 1307 -----QTTVINRATNS--LIPGSALLTKVDKDGKTLAGAEFSVRDRHNNVIRG----- 1353
QY 704 RQTTQTRQDMTTAGKALLSKSNLRQGEKQKFLQE---AQFPLGRVLDLKAIAKAEKALV 760
Db 1354 -----YEKUTTNDQOQIEATNLRPGDYQ--FVEEKAPKQVDIDKDTIEFTIVSKQKAV 1405
QY 761 TKKATK-----NGOLLERSINKAVLAYNSAIAKANKVRLKELDLTLGL 805

Db 1406 TVTATNHLKGGVTLTKTDDIDGTALAGAIPIVDANDEKKVIRENVK-----TG- 1455
Qy 806 VEGKPLAQATWVQGVY-LLKPLPLPYIYGLN---VYFKSGKLIYALDMSDITIGBQ 861
Db 1456 --ADGKVTVKOLEPGTYKFIETAP-KDVLNANPIETIDKSQQSFATVATNLSLKTGE 1512
Qy 862 KDAGNPILNVD--DNESGYHALAVATLADYEGLDIKILSLKLSQLSIRQVPTAAVHR 919
Db 1513 VE-----LLKVDGFGDKPLKG-AVFKIVDVNNNDVRTDLTTDADGKTAKDKL-----R 1560
Qy 920 AGIFOAIONAAAEQQLPKP 940
Db 1561 PGTXYFIETAPVHLRAEP 1581

RESULT 78

US-09-815-242-5639
; Sequence 5639, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5639
; LENGTH: 2086
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5639

Query Match 3.6%; Score 185; DB 9; Length 2086;
Best Local Similarity 18.3%; Pred. No. 0.043;
Matches 201; Conservative 150; Mismatches 417; Indels 328; Gaps 41;

Qy 14 TVSVYTHNQVPSLVKEPIKQTAQSSISGADYAESGSKLKI-NETSGPVDVTDL 72
Db 595 TPVTVTTSFV-----TAAANQVNSAK-QELNGDERLVRKQNAITDALITQL 642
Qy 73 FSDKRTPPE--KIKDNLKGPREQELKAVTENTESEKQITSGSQLEQSKESLSLNKTVPS 130
Db 643 -----NTPQAKLKGQVQGNLEDIQTVTQGNALNNAMKG-----LRDSIANETTVA 692
Qy 131 TSNWEIC-----DPIKGTNLVGLSKGV----- 154
Db 693 SQNYTDASPNNOSTYNSAVNAKGIINQTNPNPTMDTSAITQATTQVNNAKGLNGAENLR 752
Qy 155 -----EKLSTDLHLVPSQAADGTQLIQVASFAPTPDKKTAIAEYTSRAGENGESQL 207

Db 753 NAQTAQNLTLSHL-----TNQKSAISSQIDRAGHVSEVTA 792
Qy 208 DVDGKEIINEGEVFNYSLLKVTIPTGYKHIGODAFVNDKNIAEVLNLPESLETISDYAPA 267
Db 793 KNAATELNTMGNEQAIDHQNTVKQGVNFTDAD---KAKRDAVTNAVSRAETILNKTOG 849
Qy 268 HLAIKQDIDLPNLKAIAGELAFFDQITGKLSPLQMLRABAFKSNHIKTIETFRGSLK 327
Db 850 ANTSKQ-----DVEAAIQNVTSKALNGDQNV-----TNAKNTAKHALNNLT 892
Qy 328 VIGBASPOD--NDLSQLMPLDGLKISEAFTGPNDDHYNNRVVLTWKSKNPSGLATE 385
Db 893 SINNAQKRDLTTKIDQATTAGVAVSNTGTQLTAMANLQNGI-----NDKANTLASE 946
Qy 386 NTYVNPDKSLWQESPEIDYTKWLEDEFTYQKNSVTGFSNKGLOKQVGRKNKLEIPKOHNGV 445
Db 947 N-YHDADSD-----KKTAYT-----QAVTNAENI----- 969
Qy 446 TITIGDNAPRVDPQNTLKKYDLBEVKLPSTTRKIGAPAFQSNNLKSFASDDLEIK 505
Db 970 ----LNKNSGSNLD---KAAVENALSQV-----TNAKGALNGHNLEQAK 1007
Qy 506 EGAFMNNRIETLE-----LKDKLVTIGDAAPHINHIIVLPESVQEIERSAFRQNGANN 560
Db 1008 SNA--NTTINGLOHLTTAQDKLQVQQAQNV-----AGVDTVKSSANTLNGA-- 1054
Qy 561 LIFMGSKVTLGEMAFILSNRLEHLDLSEKQKLTETIPVQAFSDNALKEVLLPASLKTIRBE 620
Db 1055 ---MGLTRNSIQDNTATNNGQNYLDATESN----- 1081
Qy 621 AFKQHLKQLEVASALSIAFNALDDNDGDQFQPNKVUVKTHNSYALADGEHFIIVDPDK 680
Db 1082 --KTYNNAVDSANGV-----INATSNPNMDANAIQATQVTTSTKNAL-DGTHNLTOAKQ 1134
Qy 681 LSSTIVD---LEKILK-----LIEGLD----- 699
Db 1135 TATNAIDGATNLNKAQDKALKAOVTSQAORVANVTSIQOTANELNTAMGOLQHGIDDENAT 1194
Qy 700 ----YSTLRQTQTFRDMTTACKALLSK---SNLRQGEKQKFLQEA---QFFLGRVD 747
Db 1195 KQTKYRDAEQSKKTAYDQAVAAAKAILNKQTSNSDKAAVDRAALQVSTSTKDALNGDAK 1254
Qy 748 LDKAIAKAEKALVTYKATNGQ---LLERSINKAVLAYNNSAIKANVYKREKELDILLTGL 805
Db 1255 LAEAKAAKQNLGTLNHTNAQRTALEGQINQOATTVDGVNTV-KTNANTLDGAMNSLQGS 1313
Qy 806 VEGKPLAQATWVQGVVLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDITIGSQKDAY 865
Db 1314 INDK-----DATLRNQNYL-----DADESKKNATQAVTAAEGILNKQGT 1353
Qy 866 GNPILNVDENEGYHALAVATLADYEGLDIKILSLKLSQLTSIRQVPTAAVHRAGIFQA 925
Db 1354 GN-TSKADVDN---ALNTVTTRAKAALNGAENLNTKTSATNTINGLPLNLQLOK----- 1403
Qy 926 IONAAAAEABQLPKPGTH--SEKSSSESANSGKQGLQSNPKTNRGHSAIILPRTGSKGS 983
Db 1404 -DNLKHQVEQANVAGVNGVKDKGKNTLNTAMGALRTSIQDNTTKTSQNYLDASDSNKN 1462
Qy 984 F-----VYGILGVTS 993
Db 1463 YNTAVNNANGVINVTN 1478

RESULT 79

US-10-080-608A-11
; Sequence 11, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES


```

Db      459 QMEEMKTRHKGEMENALSYNITVNEQIKLMNVAINELNLIKLODTSQKEKLKEELGL 518
QY      407 WLEEDFTYQKN-----SVTGFNSKGLQVKR---NKNLEIPKQHNGVITE-----IGD 452
Db      519 ILEEKALQROLEDVLBELSPSREQIQRARQTIAEQESKLNBAHKSLSVTEDLKAEIVSA 578
QY      453 NAFR-----NVDQNKTLKYDLEEVKLPSTIRKIGAPAFQSNILKS-----FEASDDLE 502
Db      579 SESRKELEKHAEBVTNYKIKLEMLEKEKNAVLDRMAESQEAELERLRTQLLFSHEEELS 638
QY      503 EKEGAFMNNRIETLEKDKLVITIGDAAFHINHIAIVLPESVQEIGRSAPRONG-----A 558
Db      639 KLKEDLEIHRINIEKLKDNL-----GIHYKQOIDGLQNEMSQKIETMQFEKDNLITKQ 692
QY      559 NNLIIPMGSKVKTGEMAFPSNRLHDLSEQKLTETIPVQAFPSDNALKEVLLPASLKTIIR 618
Db      693 NQLILEISKLDL-QQSLVNSKSEMTL-----QINE-----LQKEIETLR 732
QY      619 EAPFKNNHKKO-LEVASALSHIAFNALDDNDGDFDNKVVVVKTHNSYALADGHEFTVD 677
Db      733 QEEKEGTLEQVQELQKTELLEKQKKEKENDLQ-----EKPAQLEAENSILK 781
QY      678 PDKLSSITVDLEKILKIEGLDYSTLRQTOTQPRDMMTAGKALLSKNLRQEGKQKFLQ 737
Db      782 DEKKT-----LEDMLKI-----HTPVQSEERLIFLDSIKS-----KSKDSVWEKEIEILIE 827
QY      738 EAQFFLGR-VLDKAIKAEKALVTYKATNGQL-----LERSINKA 778
Db      828 ENEDLKQOCIQLNBEIEKQRNTF--SFAEKNFEVNYQELQEBYACLLKVKDDLEDLSKNKQ 885
QY      779 VLAYNNSAIKKAIVKLEKELD 801
Db      886 ELEY-----KSKKALNEELHL 902

```

RESULT 81

```

US-10-171-311-2
; Sequence 2, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chan, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Moohan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hobbs, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-2

```

```

Query Match      3.6%; Score 185; DB 14; Length 3907;
Best Local Similarity 19.5%; Pred. No. 0.1;
Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

```

```

QY      7 TVALTLTVTVTHNQVFSIAKVEPIKQT--QASSSISGADYAESGSKKLKINETSGP 64
Db      82 TIMTILSGEITSHEQG-FSVELESEISTTADDCSSVNGCSFVMRTGKPTNLREBEFG 140
QY      65 VDDTVTDL-PSDKRTTPEKIDKNLAKGPRQELKAVENTESEKQIITSGSO-LEQSKESL 122
Db      141 VDDSYSGQAQDSPTLHEMSELA--GKQHEIBELNRELE-EMRVITYGTGLQQLQEF 197
QY      123 SLNK-----TVPSTSNWEICDFTIKGNLTVGLSGVGEKLSQTDHLVLPQAAQDTQLIQ 177
Db      198 AAIKQDGIITQLTANLQAR-REKDETM-----REFLELTREQSKQLQIQFOQLQASSETLR 252
QY      178 VASFAPTP-----DKKTAIAEYTSAGENGELISQLDVQKEIINEGEVFN---SVLLAKVTV 230
Db      253 NSTHSTAADLLQAKQQLTHQOOLEBQDHLE-DYQKKK-----EFTMQISFLQEKIK 306
QY      231 IPTYKGIHQDAPVDNKNIAEVNLPESLETISDYAFALHALKQIDLPNL-----KAIGEL 286
Db      307 V-----YEMEQDKKVENSKEBIEQKETIIEBNTKIIEBKKTLELKDCLTTADKLLGEL 362
QY      287 AFFNQITG-----KLSLP-----RQMLRAERAFKSNHIKTIETPRGNS 325
Db      363 ---OEQIVQKNQEIKNMKLELTNSKQEROSSEIKQLMGTVBELQRNH-KDSQF-----414
QY      326 LKVIGEASPODNDLSQMLPD---GLEKIESEAFNGPDHNNRVVLTWK-----374
Db      415 -----ETDIVQMEQETQKLSQRAEL-----DEMYQQQIVQMKEILRQMA 458
QY      375 -----SGKNPSGLAT-ENTYVNPDK-----SLWQESPEIDYTK 406
Db      459 QMEEMKTRHKGEMENALSYNITVNEQIKLMNVAINELNLIKLODTSQKEKLKEELGL 518
QY      407 WLEEDFTYQKN-----SVTGFNSKGLQVKR---NKNLEIPKQHNGVITE-----IGD 452
Db      519 ILEEKALQROLEDVLBELSPSREQIQRARQTIAEQESKLNBAHKSLSVTEDLKAEIVSA 578
QY      453 NAFR-----NVDQNKTLKYDLEEVKLPSTIRKIGAPAFQSNILKS-----FEASDDLE 502
Db      579 SESRKELEKHAEBVTNYKIKLEMLEKEKNAVLDRMAESQEAELERLRTQLLFSHEEELS 638
QY      503 EKEGAFMNNRIETLEKDKLVITIGDAAFHINHIAIVLPESVQEIGRSAPRONG-----A 558
Db      639 KLKEDLEIHRINIEKLKDNL-----GIHYKQOIDGLQNEMSQKIETMQFEKDNLITKQ 692
QY      559 NNLIIPMGSKVKTGEMAFPSNRLHDLSEQKLTETIPVQAFPSDNALKEVLLPASLKTIIR 618
Db      693 NQLILEISKLDL-QQSLVNSKSEMTL-----QINE-----LQKEIETLR 732
QY      619 EAPFKNNHKKO-LEVASALSHIAFNALDDNDGDFDNKVVVVKTHNSYALADGHEFTVD 677
Db      733 QEEKEGTLEQVQELQKTELLEKQKKEKENDLQ-----EKPAQLEAENSILK 781
QY      678 PDKLSSITVDLEKILKIEGLDYSTLRQTOTQPRDMMTAGKALLSKNLRQEGKQKFLQ 737
Db      782 DEKKT-----LEDMLKI-----HTPVQSEERLIFLDSIKS-----KSKDSVWEKEIEILIE 827
QY      738 EAQFFLGR-VLDKAIKAEKALVTYKATNGQL-----LERSINKA 778
Db      828 ENEDLKQOCIQLNBEIEKQRNTF--SFAEKNFEVNYQELQEBYACLLKVKDDLEDLSKNKQ 885
QY      779 VLAYNNSAIKKAIVKLEKELD 801
Db      886 ELEY-----KSKKALNEELHL 902

```

RESULT 82

```

US-10-756-149-5739
; Sequence 5739, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND

```



```

QY 503 EIKGAFMNNRIETLEKDLKLVITGDAAPHINHIYAIVLPESVQIGRSAPRQNG----A 558
DB 651 KUKEDLEIHRINIEKLKDNL-----GIHYKQIDGLQNEMSQKIETMQFEKQMLITKY 704
QY 559 NNLIIPMGSKVKTIGEMAFLSNRLEHLDLSEKQQLTEIPVQAFSDNALKEVLLPASLKTI 618
DB 705 NQLILEISKXKDL-QQSLVNSKSEMTL-----QINE-----LQKEIEILR 744
QY 619 BEAFKXNHLKQ-LEVASALSHIAPNALDDNDGDEQFNKVVVKTTHNSYALADGHEFIVD 677
DB 745 QBEKEGTLEQVQVQLKTELLEKQMKENDLQ-----EKFAQLEAENSILK 793
QY 678 PDKLSSTIVDLKILKLEGLDYSLTROTQTQPRDMMTAGKALLSKNLRQGEKQKFLQ 737
DB 794 DEKKT-----LEDMLKI-----HTPVSQERLIPLDSIKS-----KSDSVWEKEIEILIE 839
QY 738 EAQFFLGR-VLDLKAIAKAELVKKATKNGQL-----LERSINKA 778
DB 840 ENEDLKQCIQLNEIEKQRNTF--SPAENKFEVNYQELQEBYACLLKVKODLSDSKNKQ 897
QY 779 VLAYNNSAIKXANVKRLKELD 801
DB 898 ELEY-----KSKLKALNEELHL 914

RESULT 84
US-10-408-765A-1839
; Sequence 1839, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Boin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1839
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1839

Query Match 3.6%; Score 185; DB 16; Length 3911;
Best Local Similarity 19.5%; Pred. No. 0.1;
Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

QY 7 TVALITTVSVTVNQVPSLVKEPIKLOT--QASSISGADYAESGSKLKNETSGP 64
DB 94 TIMRTLHSGEITSHEQG-FSVELESEISTADCSSEVNGCSFVMTGKPTNLLREBEFG 152
QY 65 VDDTVTDL-FSDKRTTPEKIKDLAKPREQBLKAVTENTESEKQITGSGQ-LEQSKESL 122
DB 153 VDDSVSEGAQDSPTHELMMESELA--GKHIEBELNRELE-EMRVYTGTEGLQQLQBE 209
QY 123 SLNK-----TVPSTGNWEICDFTKNTLVGLSGVKEKLSQTDHLVLPQADGTQLIQ 177
DB 210 AAIKQDGIITQLTANLQAR-REKDET-----REFLELTEQSKLQIQFQQLQASLIR 264
QY 178 VASFAFTP-----DKTAAIAYTSRAGENCEISQLVDGKEIINEGEVEN---SYLLKKVT 230
DB 265 NETHSSTAADLLQAKQQLIHTHQQLLEEQDHILLE-DYQKKK-----EDFTMQISFLQEKIK 318
QY 231 IPTGVKHIQDQAFVONKNIAYNLPESLETISDYAPAHALKQIDLPNML-----KAIGEL 286
DB 319 V-----YEMBDKKVNSNKEIEQKETIIBELNTKIEBEKKTLELKOKLTADKLGL 374

```

```

QY 287 APPDNOITG-----KLSLP-----ROLMLABEAFKSNHIKTIETFRGNS 325
DB 375 ---QEIQVQKNQEIKNMKLELTNSKQEROSSEIEIKQIMGTVEELQKRNH-KDSQF----- 426
QY 326 LKVIGEASFONDLSQLMLPD---GLEKISEAFTGNPGDDHYNRVVLWTK----- 374
DB 427 -----ETDIVQMEQETQRLQLRAEL-----DEMYQQOIVQMKQLIRQHMA 470
QY 375 -----SGKNPSGLAT-ENTVYNPDK-----SLMQESPEIDYTK 406
DB 471 QMEEMKTRHKGEMENALRSYNTVNEQDKLMNVAINELNKLQDTSQKEKLELGL 530
QY 407 WLEEDFTYQKN-----SVTGFSGNGLQKVKR---NKNLEIPKONGVTITE-----IGD 452
DB 531 ILEBKALQRLQLEDLVEELSPSREQIQRARQTIABQESKLEAHKSJSTVEDLKAETVSA 590
QY 453 NAFR-----NVDFQNKTLRKVDLEVKLPSTIRKIGAFAPQSNNLKS---FEASDDLE 502
DB 591 SESKELELKHAEAVTYKIKLEMLEKNAVLDRMAESQEAELERLRTQLLFSHEBELS 650
QY 503 EIKGAFMNNRIETLEKDLKLVITGDAAPHINHIYAIVLPESVQIGRSAPRQNG----A 558
DB 651 KUKEDLEIHRINIEKLKDNL-----GIHYKQIDGLQNEMSQKIETMQFEKQMLITKY 704
QY 559 NNLIIPMGSKVKTIGEMAFLSNRLEHLDLSEKQQLTEIPVQAFSDNALKEVLLPASLKTI 618
DB 705 NQLILEISKXKDL-QQSLVNSKSEMTL-----QINE-----LQKEIEILR 744
QY 619 BEAFKXNHLKQ-LEVASALSHIAPNALDDNDGDEQFNKVVVKTTHNSYALADGHEFIVD 677
DB 745 QBEKEGTLEQVQVQLKTELLEKQMKENDLQ-----EKFAQLEAENSILK 793
QY 678 PDKLSSTIVDLKILKLEGLDYSLTROTQTQPRDMMTAGKALLSKNLRQGEKQKFLQ 737
DB 794 DEKKT-----LEDMLKI-----HTPVSQERLIPLDSIKS-----KSDSVWEKEIEILIE 839
QY 738 EAQFFLGR-VLDLKAIAKAELVKKATKNGQL-----LERSINKA 778
DB 840 ENEDLKQCIQLNEIEKQRNTF--SPAENKFEVNYQELQEBYACLLKVKODLSDSKNKQ 897
QY 779 VLAYNNSAIKXANVKRLKELD 801
DB 898 ELEY-----KSKLKALNEELHL 914

RESULT 85
US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Giatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoarsh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-8

Query Match 3.6%; Score 185; DB 14; Length 3917;

Best Local Similarity 19.5%; Pred. No. 0.11;

Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

```
QY 7 TVALTLTVSVTHNQEVSLVKEPIKQT--QASSISGADYAESGSKLKINETS 64
DB 82 TMRTHSGEITSHQGG-FSVELESEISTTADDCSEVNGCSFVMRTGKPTNLLREBEFG 140
QY 65 VDDTVTDL-FSDKRTTPEKIDKNAKPREQELKAVTENSEKQITSGSO-LEQSKESL 122
DB 141 VDDSYSEGAQDSPTHELMWSELA--GKQHEIEELNRELE-EMRVYTGEGLOQLOEPE 197
QY 123 SLNK-----TVPSTSNWEICDFITKGNLTVGLSKSVEKLSQTDHVLVPSQAADGTQLIQ 177
DB 198 AAIKQDGIITQLTANLQOAR-REKDETM-----REFLELTEQSKLQIQFQQLQASETLR 252
QY 178 VASFAFTP-----DKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFN---SYLLKKVT 230
DB 253 NSTHSSTAADLLQAKQOILTHQOOLEEQDHLE-DYQKKK-----EDFTMQISFLQEKIK 306
QY 231 IPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFAPHLAKQIDLPDNL-----KAIGEL 286
DB 307 V-----YEMEQDKVENSKEEIOEKETIEELNWKIIEBEKKTLEKDKLTADKLLGEL 362
QY 287 AFFDNOITG-----KLSLP-----RQLMRLAERAFKSNHIKTIEPRGNS 325
DB 363 ---QEIQVQKQOEIKNMKLELNTSKQKERSSEIKQLMGTVBELQKRNH-KDSQF---- 414
QY 326 LKVICEASQDNDLSQLMLPD---GLEKIESEAFNGPGDDHYNRVVLWTK-----IGD 374
DB 415 -----ETDIVORMEQETORKLEQRAEL-----DEMYGQOIVQMKEILIRQHMA 458
QY 375 -----SGKNPESGLAT-ENTYVNPDK-----SLWQESPEIDYTK 406
DB 459 QMEEMKTRHKGEMENALRSYNTVNEDOIKNMVAINELNKLQNTNSQKELKEELGL 518
QY 407 WLEEDFTYQKN-----SVTGFNSKGLQKVKR---NKNLEIPKQHNGVITTE-----IGD 452
DB 519 ILBEKCALQOEDLVEELSFSEIQARQTIAEQESKINEAKHSLSTVEDLKAEIVSA 578
QY 453 NAFR-----NVDFQNTLRYKVDLEVKLPSTIRKIGAFQSNILKS---FEASDDLE 502
DB 579 SESRKELELKEAEVNTYKIKLEMLEKERNALDRMAESQEAELERLTQLLFSHEELS 638
QY 503 EIKGAFMNRITETLEKDLVTIGDAAFHINHIYAILVLPESVQEIERSAFRQNG-----A 558
DB 639 KLKEDLEIHRINIEKLNKUL-----GIHYKQIDGLQNEMSQKLETQFKNLITQ 692
QY 559 NNLIIFGSKVTLGEMAFISNRLHLDLSEQKLTPEIPVQAFSDNALKEVLLPASLKTIR 618
DB 693 NQLILLEISKDL-QOOLVNSKSEMTL---QINE-----LQKEIEILR 732
QY 619 EAFKQNHKLQ-LEVASALSHAFNALDNDGDGEQDNKVVKTHNSVALADGEHFIVD 677
DB 733 QBEKEKGTLEQVEQLQKTELEKQMEKENDLQ-----EKPAQLEAENSILK 781
QY 678 PKLSLSTIVDLKILKLEGLDYSLRQTTQTFQDMTTAGKALLSKSLNROGEKQKFLQ 737
DB 782 DEKKT-----LEDMLKI-----HTPVQERLIFLDSIKS---KSKOSVWEKEIEILL 827
QY 738 EAQOFFLGR-VLDLKATAKAEKALVTKKATKNGOL-----LERSINKA 778
DB 828 ENEDLQKQICQLNEEIEKQNTF--SFAEKNFEVNYQELQEEVACLLKYKDDLEDKSNKQ 885
QY 779 VLAYNNSAIIKANVKELEKLDL 801
DB 886 ELEY-----KSKLKALNEELHL 902
```

RESULT 86

US-10-171-311-6

; Sequence 6, Application US/10171311

; Publication NO. US20030087270A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Glatt, Karen

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Hoerssh, Sebastian

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY

; FILE REFERENCE: OF CERVICAL CANCER

; CURRENT APPLICATION NUMBER: US/10/171,311

; CURRENT FILING DATE: 2002-06-12

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/298,159

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/298,155

; PRIOR FILING DATE: 2001-11-14

; NUMBER OF SEQ ID NOS: 238

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 3925

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-171-311-6

Query Match

Best Local Similarity 19.5%; Pred. No. 0.11;

Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

```
QY 7 TVALTLTVSVTHNQEVSLVKEPIKQT--QASSISGADYAESGSKLKINETS 64
DB 82 TMRTHSGEITSHQGG-FSVELESEISTTADDCSEVNGCSFVMRTGKPTNLLREBEFG 140
QY 65 VDDTVTDL-FSDKRTTPEKIDKNAKPREQELKAVTENSEKQITSGSO-LEQSKESL 122
DB 141 VDDSYSEGAQDSPTHELMWSELA--GKQHEIEELNRELE-EMRVYTGEGLOQLOEPE 197
QY 123 SLNK-----TVPSTSNWEICDFITKGNLTVGLSKSVEKLSQTDHVLVPSQAADGTQLIQ 177
DB 198 AAIKQDGIITQLTANLQOAR-REKDETM-----REFLELTEQSKLQIQFQQLQASETLR 252
QY 178 VASFAFTP-----DKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFN---SYLLKKVT 230
DB 253 NSTHSSTAADLLQAKQOILTHQOOLEEQDHLE-DYQKKK-----EDFTMQISFLQEKIK 306
QY 231 IPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFAPHLAKQIDLPDNL-----KAIGEL 286
DB 307 V-----YEMEQDKVENSKEEIOEKETIEELNWKIIEBEKKTLEKDKLTADKLLGEL 362
QY 287 AFFDNOITG-----KLSLP-----RQLMRLAERAFKSNHIKTIEPRGNS 325
DB 363 ---QEIQVQKQOEIKNMKLELNTSKQKERSSEIKQLMGTVBELQKRNH-KDSQF---- 414
QY 326 LKVICEASQDNDLSQLMLPD---GLEKIESEAFNGPGDDHYNRVVLWTK-----IGD 374
DB 415 -----ETDIVORMEQETORKLEQRAEL-----DEMYGQOIVQMKEILIRQHMA 458
QY 375 -----SGKNPESGLAT-ENTYVNPDK-----SLWQESPEIDYTK 406
DB 459 QMEEMKTRHKGEMENALRSYNTVNEDOIKNMVAINELNKLQNTNSQKELKEELGL 518
QY 407 WLEEDFTYQKN-----SVTGFNSKGLQKVKR---NKNLEIPKQHNGVITTE-----IGD 452
```

Db 519 ILSEKALQORQLEDLVEELSPREQIQARQTIARQESKLNBAHKSLSLTVEDLKAEIVSA 578
 QY 453 NAFR-----NVDFONTKRYDLEEVKLPSTIRKIGAFAPQSNLKS---FEASDDLE 502
 Db 579 SESSKELELKHAEVNTYKIKLEMLEKKNVLDRAESQBAELERLRTQLLFSHEEELS 638
 QY 503 EKEGAPMNNRIETLELKDCLVTIGDAPHINHIYAIVLPESVQIGRSAPRONG-----A 558
 Db 639 KLUKDELEHRIENIEKLKDNL-----GIHYKQOQDGLQNEMSQKIETMQFEKDLNLTQ 692
 QY 559 NNLIIFMGSKVKTGEMAFLSNRLEHLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTI 618
 Db 693 NQLILEISKLDL-QOSLVNSKEEMTL-----QINE-----LQKEIEILR 732
 QY 619 EBAFKKXHLKQ-LEVASALSHIAFNALDNDGDEQFDNKVVVKTTHNSYALADGSHFI 677
 Db 733 QBEKEKGTLEQVQELQLKTELLEKQMKENDLQ-----EKFAQLEAENSILK 781
 QY 678 PKLSSTIVDLKILKLEGLDYSLRTQTQTFRDMTTAGKALLSKSNLRQGEKQKFLQ 737
 Db 782 DEKKT-----LEDMLKI-----HTPVQBERLIFLDSIKS-----KSDSVWEKEIEILIE 827
 QY 738 EAQFFLGR-VLDJKAIAKAEKALVTCKATKNGQL-----LERSINKA 778
 Db 828 ENEDLKQCCIQLNEIEKQRTFP--SFAEKNFVNYQELQBEYACLLKVKDDLEDSEKQK 885
 QY 779 VLAYNSAIKKNVRLKELDL 801
 Db 886 ELEY-----KSKLKALNEELHL 902

RESULT 87

US-09-815-242-12610
 ; Sequence 12610, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Chislen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Cart, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2001-02-16
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12610
 ; LENGTH: 5795
 ; TYPE: PR1
 ; ORGANISM: Staphylococcus aureus
 ; US-09-815-242-12610

Query Match 3.6%; Score 185; DB 9; Length 5795;
 Best Local Similarity 20.9%; Pred. No. 0.18;

Matches 226; Conservative 142; Mismatches 418; Indels 296; Gaps 50;
 QY 13 TTVSVVTHNQVFSLVK-----EPILKQTK-----ASSISGADYAESGK 53
 Db 3817 TTVAGV--NQETAKATELNNAMHSLQNGINDETQTKQKYLDAGPSKKSAYDQAVNAK 3874
 QY 54 SKLK-----INETSPPVDVTDLFSDDKTTPEKIKDNLAKGPRQELKAVT-----EWTE 104
 Db 3875 AILTKASQNVYDKAAVEQALQNVNSTKTALNGDAKLAKAEAKAAKQTLGLTHINNAORTA 3934
 QY 105 SEKOITSSOLE-----QSKESLSLNKTPVSTSNWEICD----- 138
 Db 3935 LDNEITQATNVEGNTVTKAKAQQLDGMQQLETSIRDKDTTLQSQNYQDADAKRTAYSQ 3994
 QY 139 -----PITK-----GNTLVGLSKSGVEKLSQTDHLVLPSSQAADGTQLI-----QVASPAFT 184
 Db 3995 AVNAAATILNKTAGNT-----PKADVERAMQA--VTQANTALNGIQNLERAKQAANTAIT 4048
 QY 185 -----PDKTAIABYTSRAGENGEISOLDVDGKEIINEGEVFNISYLLKKVTIPTGYKHI 238
 Db 4049 NASDLNTKQKEALKAQVTSAGRVSAANGVEHTATLNTAMTALKHAIADKAETTKASGNVY 4108
 QY 239 QGDA-----FVONKNIAEVLN-----PESLETISDYAFHL-----ALKQIDLDPNLKAIGEL 286
 Db 4109 NADANKRQAYDEKVTAAENIVSGTPTPTLTPTSDVTNAAQTQVNTAKTQLNGHNLKVAQON 4168
 QY 287 AFFDNOITGKLSL-PRQLMRLAERAFKSN-----HIKTIEFRGNSL-----KVIGBAS 333
 Db 4169 A--NTAIDGLTSLNGPQKAKLEQVGQATTLPNVQTVRDNAQTLNTAMKGLRDSIANEAT 4226
 QY 334 PQDN-----DLSQLMLPDGLEKIESEAFYGNPGDDHNNRVVLWTKSGKNPGLATENTYV 389
 Db 4227 IKADQNYTDSQNKQTD-----YNNAVT-----AAKAIIGQTTSP-- 4262
 QY 390 NPDKSLMQESPEIDYTKWLEEDFTYKNSVTGFSN-----KGLQKVXENKNLE 437
 Db 4263 -----MNAQEIQAQ--DQVTAQOALNGQENLRTAQTNAKQHLNGLSDLTNAQKDA 4312
 QY 438 IPKHNGVTTITEIGDNAFRNVDFQNKTLRYKDYLBKVLKLPSTIRKIGAFAPQSNLKSFEA 497
 Db 4313 AKRQIEGATHVNEVTOAQNNADVLNTAM-----TNLKNQIQ-----DQNTIKQGVN 4358
 QY 498 SDDLEIEKGAFAFMNNRIETLEKDKLVIGDAAFHINHIYAIVLPESVQIGRSAPRONG 557
 Db 4359 FTDADAEAKRNAYTNAVTOAQEILNKAQGPNTAKDNVE-----TALQNVQRAKSELNG 4410
 QY 558 ANNLIIFMGSKVKTGEMAFLSNRLEHLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTI 617
 Db 4411 NQNV-----ANAKTTAK-----NTLNN-----LTSI-----NNAQK-----AALKS- 4441
 QY 618 REEAFKXHLKQLEVASALSALSHIAFNALDNDGDEQFDNKVVVKTTHNSYALADGSHFI 677
 Db 4442 -----QIEGATTVAGV-----NOVSAKATELNTAMSN----- 4468
 QY 678 PKLSSTIVDLKILKLEGLDYSLRTQTQTFRDMTTAGKALLSK-----SNLRQGEKQK 734
 Db 4469 ---LQSGIND--EAATKAAQ--KYTDADREKQTAYNDAVTAARTLLDKTAGSNDNKAAYEQ 4522
 QY 735 FLQBAQFPFLGRVLDLKAIAKAEKALVTCKATKNGQLLERSINKAVLAYNNAIKKANVCR 794
 Db 4523 ALQ-----RVNTAKTALNGDARLNEAKNTAKQOL-----ATMSHLTWA-QAN-- 4564
 QY 795 LEKELDLTLGLVEGKGLAQA--TMVQGVYLLKTPLPPEYVIGLVNVPYDKSGKLIYALDM 853
 Db 4565 LTSQIERGTTVAGVQGIQANAGTLDQAMNQLRQSIASKDATKSSBQYDQANADLQNAVNH 4624
 QY 854 SDTIGEGQKDAYGNPILNVEDNE-----GYHALAVATLADYEGLDIKTILN 900
 Db 4625 AVSDAEGIIISATNPNPNPDITINQKASQVNSAKSALNGDEKLAQAQK-----AKTDI- 4677
 QY 901 SKLSQLSIROV-PTAAVHRAGIFQAIQNAAAAEQLLPKPGT--H--SEKSSSSSESANS 955
 Db 4678 CHLSDLNNAQOTSATAEVDNAPNLAAVTSKKNKATSLNTAMGNLKHAAEKDNTKRSVNY 4737

```
QY 956 KD 957
Db 4738 TD 4739

RESULT 88
US-10-437-963-106122
; Sequence 106122, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 106122
; LENGTH: 1479
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_10596C.1.pap
US-10-437-963-106122

Query Match 3.6%; Score 184.5; DB 16; Length 1479;
Best Local Similarity 19.1%; Pred. No. 0.029;
Matches 213; Conservative 182; Mismatches 398; Indels 321; Gaps 50;

QY 27 LVKEPLKQTQASSISGADYAESGKS-----KLKINETSGPVDVT-----VTDL 72
Db 467 LUBERV--QEEAASSAEKLEAEATNSVEAYKEKINELQASLDSTTSKNQLLEQEVKDL 524

QY 73 FSDKKTTPKIKNDKAKPRELAKVAVTENTSEKQI-----TSGSQLEQSK 119
Db 525 -SDKFTHEQEAHVHE--RSLESLHTSDEAVHATQDLENELNANERKFEVE 581

QY 120 ESLSLNKTVPTSTSNWEICDFITKGNLTGLSKSG-----VEKLSQ-----TDHLVL 165
Db 582 ADLEQYRSKVSQLSDELEAYQTKAASLEAVMESASEKEKELVESLGQITEKKLELVL 641

QY 166 PSQAADGTQLIOVAS-----FAFTPKKTAIAEYTSRAGE-----NGEISOLDVDGKEIIN 216
Db 642 EYBEKTEEVLKESQSLERLOQESKVLALERSLVKMGEEKESHGHTIADNL-----QLSN 698

QY 217 EGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLN--PESLETISDYAFALHAKQI- 274
Db 699 KNDWY-----MOLESQSQAGDDHSKTRSLLSAQLHKEELE-----LNLKSLN 742

QY 275 DLPDNLKATGELAFPDNOITGKLSLPRQLMRLAERAFKSNHKTTFPRGNSUKVIGEASF 334
Db 743 DLHVESKTAESAL---ORIALETQVQELSAAEQSLKS--HLTEFESK-----LASAEK 792

QY 335 QNDLSQLMPLGLEKIES-----EAFTCNPGDDHYNNRVVL---WTKSGKNPSGLATEN 386
Db 793 KSMDEQ--ELKATECNSRLRVDELSEL--EYKKEKTSLEASLEAKQEAELSEKL 850

QY 387 TVYNPKSLQWRSPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKNLEIPKQHNGVT 446
Db 851 DQWNEEKFEELSCKATIKHLEAE-----NQVQAL--QGELESARHKLVEESDLGAL 903

QY 447 ITEIG-----DNAFRNVDFQNTLRYDLEFVKLPSTIRKIGAFAPQSNLKSFAASDDL 501
Db 904 IRETSVLDKLAEEQLEHKGALHATSKKIDLEALYQSL-----LEDTEMKLOQAGENL 959

QY 502 EEIKEGAFMNNRIETLELKDKLVITGD--AAFPHINHIYAIVLPESVQVIGRSAPRONGAN 559
Db 960 TQ-----KETEQELSEKLAAEQAAASYQAKATAAAAEVESV--KVELEAFE-----1005

QY 560 NLIFMGSKVKTIL--GEMAFISNRLEHLDLSEQQLTEIPVQAFSDNALKEVLLPASIKTIR 618
Db 1006 -----TEISTLETTIEELKTKASNAESRAEQVALVESAMMSETNQALKED--LDKALMLR 1058

QY 619 E--EAFKKNHLKQLEVASALS-----FDMKNVVKVTHHNSYALADGHEFIVDDP-----KLSSTIVDLK 690
Db 1059 ELQEQFDSHAEKEEVFTKLSAHEKTIHLETVHSGRGLHATAESKNAEQAQHEALE 1118

QY 643 ALDDNDGDEQ-----FDMKNVVKVTHHNSYALADGHEFIVDDP-----KLSSTIVDLK 690
Db 1119 TIGKKDSEVKDNLNRLAALAESEIESLTHVNEAMKQEIINAKLVKVDELQELKLSISSEKEE 1178

QY 691 ILKLEGLDYSTLRQTOTQFRDMWTAGKALLSKSNLRQGE-----731
Db 1179 VAEKV-----VVHEKTTIEHLREHSHRGLQSAESRSAAETENELREVLETVAQKEARV 1232

QY 732 --KQKFLQ---EAOFFLG-----RVDLDKAIA-----KAEKALVTK 762
Db 1233 TDLKEKLVSTENKLVINEALKGELDTKVAMPDELOEQFSSHAEKEEAELKLVH 1292

QY 763 K-----ATKNGQLLERSINKAVLAYNNSAIKANKVRLKLEKLDLL--802
Db 1293 RTISHLTEVHTRSLELSAESKNEEIESKLHEAL---EMAAQKEAEVKDLSKLDAL 1349

QY 803 -TGLVEGKGLAQATMWQGVLLKTPPLPEYIGLVNVTYFKSGKLIYALDMSDITIGEQ 861
Db 1350 ELGYVEQA--TEAAAABETHKIK-----FDEAVHKIKSLERQLAVTENK 1392

QY 862 KDAYGNPILNVDEDNQGYHALAVATLADYEGLDIKTILNSKLSOLTSIROVPTAAVHRAG 921
Db 1393 VE-----LPFTEKEN-----LVIANSKLNEELHLQ-----N 1419

QY 922 IFQATQAAAAEAQQLLPKPGTHSEKSSSESANS 955
Db 1420 KNLQVALAAV-----AEKEGSEEHIS 1444

RESULT 89
US-10-369-493-13477
; Sequence 13477, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13477
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Thermoplasma volcanium
US-10-369-493-13477

Query Match 3.6%; Score 184; DB 15; Length 891;
Best Local Similarity 19.5%; Pred. No. 0.015;
Matches 198; Conservative 159; Mismatches 335; Indels 322; Gaps 52;

QY 51 SKSKLKNETSGPVDVTVDLFSDKRTTPPKIKNDKAKPREL-----96
Db 34 AGKSSI-----VDARFALFSDKRT--RRTEDMIKKGERYMEVELYFRSEGHYSYIR 83
```

```
QY 97 -----KAVTENTSEKQ---ITSGSQLEQS---KESISLNTKVPSTSNWEICDFTKG 143
Db 84 RTIERGKSIISDAIEHDGSIITGASDVSNYKVNINIKOVFLTS-----IFVRQG 138
QY 144 NTLVGLSGVGEKLSQTHLVLPSQAADGTQLIQVASFAPTPDKKTAIAEYTSRAGE-NG 202
Db 139 EMDALVSKDPAERKKILBNIDRLRLEAGYLLK---EVIDDLTANVSDYDYLKNEQLQS 194
QY 203 EISQJVDGKEI-----INRGVFNYSYLLKKVTIPTGYKHICQDAPVNNKNAEVL 254
Db 195 KINEIDNNKQIEELESKRLTIEPI--KALEEINIKENK-----DHLNBEHLRL 244
QY 255 PESLETSIDYAF-----AHLAKOIDLP---DNLKAIGELAFF--DNOITGKLSLP 300
Db 245 NAQLETIKKYEMLAESRKSASIEWVVKGPSIEBELKRLNNAVVKRNEIIEYINLK 304
QY 301 RQLMLAERAFKSNHIKTIERGNLSKLVIGEA-----SFQDNDLSQLMPLDGLKIE 352
Db 305 KDLGSLSE-----IEGLKSDLKYDEAHRKLEDLQSPRSEFFLEKKRKEKEDLKL 355
QY 353 SEAFGTGPGDDHNNRVVLTWYKSGKGNPSGLATENTYVNPDKSLMOESPID-YTKWLEED 411
Db 356 SSL-----KEDENYQSAV-----RNIEIKKWIENEKEIERMSAFISEI 396
QY 412 FTYQK-----NSVTGFSNKGIGQYK-----RNKNLEIPKQ-----441
Db 397 LKIQITPEIINRRRAEINSSLMQIEGKIASINASIDAMRSHKMEVEENAAWLSRGVCP 456
QY 442 -----HNGVTITE-----IGDNAFRVNDPQNTK---LRKYD-----LBEVKLPSTIR 480
Db 457 VCGTHLGTEKSEDLVKHGEEASRLSEEDINKTENEIKDLBERKHQKLLDRINGKOVER 516
QY 481 KIGAFAPQSN---NLKSPASDDEBIEKEGAF-----MNRITLBLEKOKLVVIGDAA 530
Db 517 LIASYNLLSSKRAELKKEF--MDDEARLKEAHLKAEAAISQYNSIDLGLEAK---NEEW 570
QY 531 PHINHIYAIVLPESVQIEGRSAFRONGANNLI FMSGKVKTLGEMAFSLNRLHLDLSQK 590
Db 571 LKANAVISIDIENT-----RSRFE-----KNQNLNDIIRKWNIE-----607
QY 591 QLTTRIP-VQAFSDNALKEVLLPASLKTIREBAFKKNHL-----KQLEVASALSHTAFNAL 644
Db 608 ---VNPIDVESYNENSLKRI--DBELNSLRN---KNELYAKKAAMDEIQKTIHF-----655
QY 645 DNDGDGDFDNKVVVTHHNSVALADGHEFIVDPD--KLSSTVDLEKILKLEGLDYST 702
Db 656 -----KEEISKK---KGIEDSQAENVNAQLQINDDLKQLSSR---LDKI--NVQYEWKS 702
QY 703 LRQTTQTPQDMTTAGKALL---SKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKAL 759
Db 703 LH-----KVLLOQNEKLNIAVADIRKRLEKKTII-----KAIADLKR--740
QY 760 VTKKATKNG--QLLBSRNKAVLA YNNSAIKANKVRLKLEKD-----LLTGL 805
Db 741 VREAFSGKGVPAIIRKSASEPI---TNQT--RQYIQRFEIDDDVDVDQDFNITVFRGSI 795
QY 806 VEGKQPLAQATWQVQVYLLKTPLPPEYIYIGLVNVPDKSGKLIYALDMSDITIGEQKDAY 865
Db 796 AEGIDSLSGGERMAVAFALR-----VAIAQFLNKDVLSL-----VM 831
QY 866 GNPILNVDEDEGYHALAVATLADYEGLDIKTLNSKLSQLTSTIRQVPTAAVHR 919
Db 832 DEPTAFLEDDRRS-----DLANIIEYLSKDSGGIPQVIMISHRR 870
```

RESULT 90

```
US-10-282-122A-71690
; Sequence 71690, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
```

```
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71690
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-71690
```

```
Query Match 3.6%; Score 184; DB 15; Length 1665;
Best Local Similarity 19.4%; Pred. No. 0.036;
Matches 210; Conservative 186; Mismatches 430; Indels 258; Gaps 50;

QY 16 SVVTHNOEVFLVKEPILKOTQASSISGADYAESGSKLKNETSGPVDVTV--TDLF 73
Db 365 NVINNTTEATEERKQVALTKLN-----KYNEKKEIDNLISSSDIA 406
QY 74 SDKRTTPEKIDNLAQGPQELK-----AVTEN-----TESEKQITSGSQLEQS 118
Db 407 PLKESSIAQINSINVRATKAAAKQAIAEALTDRKVFIDSHYDATQEKDVAMAKVTSEA 466
QY 119 KESISLNTKVPSTSNWEI-----CDFITKGNLTVLGSLSGVGEKLSQTHL 163
Db 467 NKAKAL--IDQSTSNNDVDQAQTNGINIINSIDADVTKKAN-----AKKAIEQAEEAKKA 519
QY 164 VLPQQAADGTQLIQVASFAPTPDK---KTAIAEYTSRAG---ENGEISOLDVDGKEII 215
Db 520 LI-NQNSDATQEEKDAAIQRTVDEVRNADRLIDQSTNNDGVDEVOAHSSINNIQPIV 578
QY 216 NEGVRFNYSYLLKKVTIPTGYKHICQDAPVNNKNAEVLNPESELETISDYAFAPHL--ALUK 273
Db 579 KKSDA-----KQAIDTAV--LNQKSLVNNNN-----EATQEKDVALAKIDEAAKQ 622
QY 274 ----IDLDPNLAIGELAFPDNQLTKLS--LPQLMLAERAFKSNHIKTIERGNLSK 327
Db 623 AKAAIDAATTNNAVDEAT---NNNTTIISGLPDTVKKAAAR-----KAIDDAATAKK 672
QY 328 VIGESAFQDNDLSQLMLPDGLKIE-----SEAFGTGPGDDHNNRVVLTWYKSGK 377
Db 673 ---EAINNTSDATQEEKDAALAKVAAVTAAKQAITQATTNDNDVQEQN-----718
```


Db 663 DAQALSATNGNHLSDKSELKQESD-----MEDIDGANLCADPGVVVPALHGBTTSSDLAD 717
Qy 671 GEHFIV-----DPDKLSSTIVDELEKILKLEIGLDYSTLRQTTQTPR 712
Db 718 NDGAENVSTPACDGLASSGAPTGNDSKNSAAAQAQVEEDVPQDDAQVEEDVPSQDD 777
Qy 713 DMTTAGKA--LLSKNLNRQGEKQFLQEAQFLGRVLDLKATAKAELAVTKATKNGQL 770
Db 778 DNPADGAPGEICSNANAFSTSSCAVEYV-----QDIASVTVDIHDKNDND--- 827
Qy 771 LERSINKAVLAYNNSAIIKANVRLKLEKELDLTLGLVEGKGPLAQATMVQGVYLLTPLPL 830
Db 828 -DENINTDITG-NHSEPK-----LETNVD-----NEURGDI-QVKKPYVYLMKVP--- 870
Qy 831 PEYIYGLNVYFDKSKGLIYALDMSDTIGEGQKDAYGNPILNVDENEGVHALAV----- 884
Db 871 -----RPMSESHWEKIQDAQ-----ICDELTKRDAINVLQRKKK 906
Qy 885 ATLADY-----EGLDIKTLNLSKLSQLTSSIRQVPTAAVHRAGIQAIQNAAAAEQ 935
Db 907 ALCDYRLEKLEAAROEERGAHTAGDKNDLNSVQSM-IGRMNTRANSIQEIDDMTAMKEK 965
Qy 936 LLPKPGTHSEKSSSESANSK-DRGLQSNPKTNRGRHSAILPRTGSK 981
Db 966 II-----AHESI5LKEKELLQDIKELKAKQKQLSSNWGSK 1001

RESULT 95

US-09-815-242-11391
; Sequence 11391, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kazi L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11391
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11391

Query Match 3.6%; Score 182.5; DB 9; Length 944;
Best Local Similarity 19.2%; Pred. No. 0.02;
Matches 201; Conservative 148; Mismatches 327; Indels 371; Gaps 49;

Qy 6 KTVALTITVSVVTHNQ--EVFSLVKEPIKLTQASSISGADYAESGSKLKNETSG 63

Db 29 KDIGLELKTNSKMTPEQAGKLYKIVDGIKEIQANQP---AKNPQDNKDDLATAVASK 85
Qy 64 PVDDTVTDLPDSKKT---TPEKJKDN-----LAGPREQELKAVTEN-----TESBKQ 108
Db 86 SLNKKVSKTPKKEETKSPQPKPKTKKKEKGAFTPIAKKKGGEIIVNTFENQTPPTENTPK 145
Qy 109 ITSGSQLEQSKESLSLNKTVSTNWEICDPIITKGNITVLGSKSGVEKLSQTDHLVLP SQ 168
Db 146 VVSHSQIEKAKQKIQ-----EIQKSRREALNKLTSN----- 176
Qy 169 AADGTQLIQIVASFAFTPKKTAIAEYTSRAGE-NGEISQLDVDGKEIINEGEVENSYLK 227
Db 177 -----ANNASNANNAKKEISEVKKQEQEIKRH-----E 204
Qy 228 KVTIPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFAHALKQIDLDPNLKAIGELA 287
Db 205 NIKERTGFRVIKRNDENEVE--SENSVTESKKPTQSAAAIPEDIKK-----EWQ 251
Qy 288 PFDNQITGKLSLPLQMLRAERAPKSNHIKTI-----EPRGNSLKVTIGEASPODNDLSQ 341
Db 252 EKDQEAACKAKKPSK--PKATPTAKNNKSHKIDFSDARDPKGNDI-----YDDETD 301
Qy 342 LML-----PDGLEKIESEAFNGDDHNNRVVLTGSKGNPSGLATENTYVNPDKSLW 396
Db 302 ILLFDLHEQDNFKNEEBE---KSIQRNINDRVV---QRKNP----- 337
Qy 397 QESPEIDYTKWLEED-FTYQKNSVTGFSNKGLOKQKRNKNLEIPKQNGVTVITIGDNAP 455
Db 338 -----WMNESGIKROSKKRAFRNDNSQVIOSTTAIPDEE---VRVYFAQKAN 383
Qy 456 RNV-----DFQNTLRKYDLEEVKLPSTIRKIGAFQFQSNNLKSPFAS 498
Db 384 LNLADVIKTLFNLGLMVTNDFDKDSIEILAEEFHLSEISVQ-----NTLEEPEVE 434
Qy 499 DDLBEIKE-----GAPMNNRIETLE-LKDKLVTIGDAAPHINHYAVLPES--- 544
Db 435 EVLEGVKKERPPVVTIMGHVDHGKTSLLDKIRDKRVAHTEAGGITQHGATGMVKNKDW 494
Qy 545 ---VQEIQRSAF---RONGAN---NLIFM---GSKVKTIGEM-----AFLSNRL 581
Db 495 SFIDTPGHEAFSOMRNRGAQVTDIAVIVIAADGCVKQQTIEALEHAKAANVPVIFAMNM 554
Qy 582 EHLDSLSEOKQLTE-----IPVQAFS-----DNALKEVLLPASLKTREE 620
Db 555 DKPNVNPDKLAKCAELGYNPVDWGEHEFTIPVSAKTGDGIDNLETLIQAGIM----- 609
Qy 621 AFKNHLKQLBVASALSALHIAFNALDDNDGDQFNDKVVVKTTHNSYALADGEHFTVD--- 677
Db 610 -----ELKAIEGSAARAVVLEGSVEKGRG-----AVATVIVOSGTLVSQDSFPAETAF 657
Qy 678 -----PDKLSSTIVDLE-KILKLEIGLDYSTLRQTTQTPRDMTTAGKALLSKSN----- 726
Db 658 GKVRTMTDDQKSIQNLKPSMVALITGLS-----EVPAGSVLIGVENDSIA 704
Qy 727 -LROGEKQKFLQEAQFFLGRVLDLKATAKAKA-----LVTKKATKN-----G 768
Db 705 RLQAKGRATYLRQ-----KALSKSTKVFDSELMVANKELKNIPVVIKADTQ 753
Qy 769 QLLERSINKAVLAYNNSAIIKANVCR-----LEKELDL-----TGLVEG 808
Db 754 SL--EAIKNSLLELNNEEVAIQVHSGVGGITENDLSVSSSEHAVILGFNIRPTGNVK 811
Qy 809 KGPLAQATMVQGVYLLKTPPLPPEYIYGLNVYFDKSKGLIYALDMSDTIGEGQKDAYG-- 866
Db 812 KA-----KEYNVSIKTY-----TVIVAL-----IEBMSRLILGLM 841
Qy 867 NPILNVD-----EDNEGYHALAVATLA 888
Db 842 SPIIEEHTGOAEVRETFNIPKVGTTIA 868

RESULT 96

[illegible]

Qy	288	FFDNQITKGLSLPQLMRLAERAKSNHKT	-----EPRGNLSKVIGSEASQDDNLSQ	34	
Dd	252	EKDQEAkakpsk--pkatpaaknshki	DFSDARDFKNDI-----YDETDE	301	
Qy	342	LML-----PDGLEKIESEATGPNRVL	WTSGKNPSGLATENTYNPKSLW	396	
Dd	302	ILLFDLHEODNFNKEEBE	---KEIRQNINDRVRV---QRKNP-	337	
Qy	397	QESPEIDYTKWLEED-PTYOKNSVTGS	NKGLOKVRKNKLEIPKOHNGVTITE	455	
Dd	338	-----WMNESGIKRQSKKKRAFRD	NSORVQS-TTALPEE---VRVTFPAKAN	383	
Qy	456	RNV-----DFQNKTLRKYLDEBVKLP	STIRKIGAFAPFOSNNLKSPFEAS	498	
Dd	384	LNLADVIKTLFNGLMVTKNDFLDKDS	IELAEHFLEISVQ-----NTLEEFEV	434	
Qy	499	DLLEEIKE-----GAPMNRIETLE-L	KDKLVITGDAAAHNIHVAIVPES---	544	
Dd	435	EVLEGVCKERPVPVVTINGHVHGKT	SLLDKIRDKRVHAETEAGGITQHIGAY	MVERKDW 494	
Qy	545	--VQEIGRSAP-----NLIFM----	GSKVKTUGEM-----AFLSNRL	581	
Dd	495	SFDTTGFHEAFSQWRNRGAQVTDIA	VIVIAADGVKQOTIEALEHAKAANPV	FANMKM 554	
Qy	582	EHLDSLSEQOLTE-----IPVOAFS-	----DNALKEVLLPASLKTREE	620	
Dd	555	DKPNVPDKLKAECALGYNPVDWGGB	HEPIPVSAKTGDDGIDNLTILIQAGIM	---	609
Qy	621	AFKNHLKQLEVASALSIIAFNALD	NDGDGEQDPNKVVVTHINSVALADGE	HFIVD---6777	
Dd	610	-----ELKATEBSARAVLEGSVERG	RG-----AVATVIQSGTLSVGDSFF	AETAFA 657	
Qy	678	-----PKLSSTIVOLE-KILKILEGL	DYSTLRQTQTQFRDMTTAGKALLSKN-	---756	
Dd	658	GKVRTMTDDQKSIQNLKPSNVALIT	TGLS-----EVPAGSVLIGVENDS	IA 704	
Qy	727	-LRQGEKQFLOBQAQPLGRVDLDKA	IAKAEGA-----LVTKKATKN-----	G 768	
Dd	705	RLQAQKRAATVLRQ-----KALS	SKTSKVSFDBELSEMANKELKNIPV	VIKADTQG 753	
Qy	769	OLLERSINKAVLAYNNSAIKANVR-	-----LEKELDLL-----TGLVEG	808	
Dd	754	SL--EAIKNSLLELNNEEVAIQVHS	GVGGITENDLSVSSSEHAVILGFNIR	PRTGNVK 811	
Qy	809	KGPLAQATMVQGVLLKTPPLPEYLI	GLNVYFDKSGKLIIAYALDMSDTTGEG	KDAYG--866	
Dd	812	KA-----KEYNVSIXTY-----	TVIYAL-----TEMRSLLLGLM	841	
Qy	867	NFILNVD-----EDNEGTHALAVATLA	888		
Dd	842	SPIIEBHGTQAEVRETFNPKVGTIA	868		
 RESULT 97 US-09-815-242-10903 ; Sequence 10903, Application US/09815242 ; Patent No. US20020061569A1 ; GENERAL INFORMATION: ; APPLICANT: Haseelbeck, Robert ; APPLICANT: Ohlsen, Kari L. ; APPLICANT: Zyskind, Judith W. ; APPLICANT: Wall, Daniel ; APPLICANT: Trawick, John D. ; APPLICANT: Carr, Grant J. ; APPLICANT: Yamamoto, Robert T. ; APPLICANT: Xu, H. Howard ; TITLE OF INVENTION: Identification of Essential Genes in ; FILE REFERENCE: ELITPA.011A ; CURRENT APPLICATION NUMBER: US/09/815,242 ; PRIORITY FILING DATE: 2001-03-21 ; PRIOR APPLICATION NUMBER: 60/191,078 ; PRIOR FILING DATE: 2000-03-21					

```

/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10903
/ LENGTH: 1192
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
US-09-815-242-10903

Query Match      3.6%; Score 182; DB 9; Length 1192;
Best Local Similarity 19.0%; Pred. No. 0.03;
Matches 218; Conservative 177; Mismatches 431; Indels 324; Gaps 51;

QY 46 DYAES- - - - -GSKLKINETSPPVDDTVTLFSDK- - - - -RT 78
Db 94 DYSEISVTRRLKRTGESDFINKACRLKD-IQDLFMDGSLGKESFSIISQGVKVAIFNS 152
QY 79 TPBK- - - - -IKONLAKGPREBELKAVTNT- - - - -ESEKQITS-QSOLEQS 118
Db 153 KPEDRRGIFEEAAGVLKQKQKKAQKLFETEDNLSRVQDIYIELEDQLVPLAAQADAA 212
QY 119 KESLSL- - - - -NKTVPSTN- - - - -WEICDFITKNTLVGLSKSGVEKLSQTDHLVL 165
Db 213 KYLAKBELTEIDVNLVTETIQEAKAIWE- - - - -TKQELTAIE- - - - -EKLASQKVH 263
QY 166 PSQAADGTOLIOVASPAFTPKKTAIAEYTSRAGENGEISQLDVDGKEIINEGEVFNFSYL 225
Db 264 DLEG- - - - -KLVLRS- - - - -KRNRLDE- - - - -QIETEQQQLQVTEALKAQE 301
QY 226 LKVTIPYQKIGQDAFVDNKNIAE- - - - -VNLPESETISDVAFAPHLAKQIDLPNL 280
Db 302 GQKNVLIBSKHSTQTSASEYBETLAETAKIVRYREELQTLTETAIAEKTAQORQ- - - - -TL 356
QY 281 KAIGELAPFDNQITGKLSLPROLMR- - - - -LAERAFKSNHKTIEPR- - - - -GNS 325
Db 357 KEALATKDVKEYSKS- - - - -KELMBELRSQYVEVMEQANTANDKYLERYQOQETAKNQ 414
QY 326 LKVIQASPDNDLSQJLMDGLEKIESEAFGTG- - - - -NPGDDHYNRNVLMVTKSGKNPSGLA 383
Db 415 QSLAKHEALEEQMVEALAMKETLEKEQKQVAKQGLQEQLEEVYALKATLEAK- - - - -465
QY 384 TENYVNPDKSQQWESPEIDYTKWLEEDFTYQKNSTVQFSNKGLOKQVKNQ- - - - -435
Db 466 -RERLAQRNDMYQAMNOVQQAQKQKSLQEIQENYAGF-YQGVKAVLRHKNQLTGIVGA 523
QY 436 - - - - -LEIPKOHNGVTITEGDNAFRNVPFNQNTLKYDLEVK- - - - -LP-STIRK 481
Db 524 VAEILVBPKEYTLAIEALG-GAAQHVVENEKGKRGITFLKQHSGRATFLPLTIKPP 582
QY 482 IGAFAPQSNLKSFEA- - - - -SDDLBEIKE- - - - -GAFMNNRIETLEKDKLVITGDAAFINH 535
Db 583 RSVSAMVQNRLAGAPGVGIASELVRYPEQVTVIQLNLGVITLAADLTSANQLAKLVNY 642
QY 536 IYAIIV-LPESVQEIGRSAPRQNGANNLIPMGSKVKTILGEMAFPLNRLEHLDLSQKQITE 594
Db 643 QYRVVSLGDDVMPGGS- - - - -MTGANKRGNGSLFSPAQELQITTEQMTQLE- - - - -TQLRSVE 698
QY 595 IPVQAFSDNALKEVLLPASLKTIRBEAFKKNHLKQLEVASALSALHAFNALDDNDGDQFDP 654
Db 699 QEVALS- - - - -QEVKTATRAEMLRASQGNRLKQOEI- - - - -D 733
QY 655 NKVVVTHNSVALADGEHFIVDPPKLGSTIVDLKILKLEGG- - - - -LDYSTLRQ 705

```

Qy	196	SRAGEICEISOLDVDGKEIINEGEVFNYSLLKKVTITPTGYKHIGQDAFVNKNIAEVLNP	255
Db	924	EEVKE--EVATTLI---ETVEQAEBSKANTITET-----FENLEENAVESNENVA-----	968
Qy	256	ESLETTISDYAPAHALAKQIDLPDNLKAIGELAFPDNQITQKLSIPLQOLMRLAERAPKS--	313
Db	969	ENLEKNETVF-NVTLDKVE--ETVEISGE-----SLENNEM--DKAPFSEI	1010
Qy	314	-NHIKTIE-----FRGNSLKVIGEASFOONDLSQMLPDGLEKIESEAFITGNPGGDHY	365
Db	1011	FDNVKGIQENLLTGMRFSIETSIIVIO--SEEKVDLNNVWSSILLNIEN-----	1057
Qy	366	NNRVLWTGSKGPKSLATENTYNNPKSLWQSPRE-IDYTKULEEDFTYQKNSVTGFSN	424
Db	1058	-----MKEGLLNKLENISSTEGVQETVTEHVQNVVYDVPAMKQDFLGILN	1105
Qy	425	K--GLQKVKNKNLEIPKQHNGVTITBIGNAPRNVDFONKTLRKYD-----LEEV	473
Db	1106	EAGLKEMPNLEEDVPKSESDVITVEIKBPVQK-EVEKETVSIITEEMENIVDVLEEE	1164
Qy	474	KLPTSTIRKIGA---FAQSNNLKSPASDD-----LEEIKEGAFMNNRIETLEKD	521
Db	1165	KEDLTQMDIDAVEESIEISSDSKEETESIKDKKEDVSVVEEQDNDMDSESVEKVLKKN	1224
Qy	522	KLVTIGDAAHHINHIYAIVPESVOEIGRSAPRONGANNLIFMGSKVKTIGEMAFISNRL	581
Db	1225	MEBELMKDAVEINDITSKLII-EETQELNEV-----EADLIKOMEKIKEL-EKALSBDK	1276
Qy	582	EHLDD-----LSEOKQLTEIPVQAFSD-----NALKEV-----	608
Db	1277	EIIDAKDDTLKUIEBEHDITTLTLDVEVELLVEEDKIEKVSULKLEEDILKEVKEIKE	1336
Qy	609	----LIP--ASLKTI-----REEAFKKNHLKQL-EVASALSHIAFNALDDNDGDEOFON	655
Db	1337	LESEILEDYKELKTIETDILEEKEIEKHDFEPEEAEIKOLEADILKEVSSLEVEE	1396
Qy	656	KVVVKTTHNSYALADGEHPI-----VDPDKLSTTIVDLEKILKUIEGLDYS	701
Db	1397	KCLEEVHE--LKEBEVHHISGDAHIKGLSEEDLLEEVDLLKGSITLMDLKGDMELGDMDK	1453
Qy	702	TLRQTTOTQPRDMTTAGKALLSKNSLRQGEKQFLQEAQFGLRGVLDLDAKAAE-----	756
Db	1454	SLESDVT-TKLGERVESLKDVSSALGMDSEQMKTRKKAQ-----RPKLEEVLLKEEVEKEE	1508
Qy	757	KALVTRKATK-----NGOLLERSINKAVLAYNNNSAIKKANVXRLESEKELDLLT	803
Db	1509	KKKITKKVRFDIKDEPKDEIVEVEKMDIEEDVEEDIEEDKVEDIDEDIDEDI	1568
Qy	804	GLVEGKGPLAQATW-----VOGVVLTKTPL--PLPEYIIGLVNVPFKSGKLIYALDMSDT	856
Db	1569	G--EDKDEVIDLIVQKEKRIEKYKAKKKLEKKEVEGVSGLKHVDVEMVKYQKIDKEVD	1626
Qy	857	IGHGOKDAYGNPILNVDEDNEGHA-----LAVATLADY-----	890
Db	1627	KEVSALESKNDVTNVLKQNDQFFSKVNFVKKYKFAAPFISAVAAFAFVYVGFFTFSL	1686
Qy	891	----EGLDIKITILNSKLQ--LTSIRQVPTAAYHRAGIFQAIQNAAAAEAEQLLKPQGTSE	945
Db	1687	FSSCVTIASSTYLLSKVDKTIINKKERPPFYSF---VFDFIKNLKHVLOQM-----KE	1735
Qy	946	KSSSSSESAN-----SKDRGLQSDNKTNR	968
Db	1736	KFKSEKNNNVIEVTNKAEEKGNQVNTNKTEK	1766

RESULT 99

US-10-415-253-2 .

00 20 113 433 2 ; Sequence 2, Application US/10415253

Sequence 2, Application 03/1041
; Publication No. US20040067236A1

GENERAL INFORMATION:

APPLICANT: Cohen, Joe

APPLICANT: Druilhe, Pierre

TITLE OF INVENTION: Immunogenic Compositions Comprising

```

; TITLE OF INVENTION: Liver Stage Malarial Antigens
; FILE REFERENCE: B45250
; CURRENT APPLICATION NUMBER: US/10/415,253
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: PCT/EP01/12349
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: EP00203724.0
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1787
; TYPE: PRT
; ORGANISM: K1 Parasite Clone
US-10-415-253-2

```

Query Match	3.6%	Score 181.5;	DB 15;	Length 1787;
Best Local Similarity	18.4%;	Pred. No. 0.057;		
Matches 204;	Conservative 205;	Mismatches 413;	Indels 289;	Gaps 47;

QY	22	QEVFSLVKBPIKQTOASSISGADYABSGSKGLKINETSGVPDDTVTDLFSDKRTTPE	81
Db	781	ESVAENVEESVAPTVEIIVAPSVVEESVAPSVESVAENATNLSDLLSNLLGGIET--E	838
QY	82	KIKONLAKPREQELKAVT-----ENTSEKQITSGSQLEQSKESLSLNKTVPSITSNW	135
Db	839	EIKDSTILNEIBEYKENVVTILLENVETTAESVTTFSNILEEIQENTITNDTIE----	892
QY	136	ICDFFITKGNITVLGLSKGVEKLSQTDHLVLPQAADGTQLIQVASFAFPDKKTAIAYT	195
Db	893	-----EKUEELHENVL--SAALENTQ-----SEEEKKEVIDVI	923
QY	196	SRAGENGESQLDVGKEIINEGEVFNLSYLLKKVTIPTGYKHIGODAFVNDKNIAEVNLP	255
Db	924	EEVKE--EVATTLI---ETVEQAEKXSANTITFEI-----FENLEENAVESNVA----	968
QY	256	ESLETISDYAFHAHLKQIDLPNLKAIGELAFFDQNTGKLSPLRQLMLABRAPKS--	313
Db	969	ENLEKLNETVF--NTVLDKVE--ETVEISGE-----SLENEM--DKAFFSEI	1010
QY	314	-NHIKTI-----FRGNSLKVIGBASQDNDLSQMLPDGLEKITESAFTGNPGDDHY	365
Db	1011	FONVKGIOENLLTGMFRSIETSIIVQ--SBEKVDFLNENVVSSILDNIE-----	1057
QY	366	NNRVVLWTKSGKNPGLATENIVNPDKSLWQESPE-IDYTKWLEEDFTYQKNSVTGFSN	424
Db	1058	-----MKGELLKNKLENISSTEGVQETVTTEHVQNVDVDPVPMKDQOFLGILN	1105
QY	425	K--GLOKVRKNKLEIPKOHNGVTIETIGNAPFRNVDFQNKTLRYD-----LEE	473
Db	1106	EAGGLKEMFPNLEDVPKSSDVITVEIKDEPVQK--EVEKETVSIIIEEMEENTVDVLEE	1164
QY	474	KLPSTIRKIGA---PAFOSNLLKSPASDD-----LEETKEGAFNNRIETLEKD	521
Db	1165	KEDLTDKMDAVEESIEISDSKEETESIKDKEKOVSLVVEEQDNDMDVESVKVLEKN	1224
QY	522	KLVITGDAAFHINHYIIVLPSVQFGRSAFRQGANLNIIMGSKVKTILGEMAFISNRL	581
Db	1225	MEBELMKDAVEINDITSKLI--BETQELNEV-----EADLIKDMEXLKEL--EKALS	1276
QY	582	EHLN-----LSQOKQLTEIPVQAFSD-----NALKEV-----	608
Db	1277	EIIDAKDDTLEKVIIEEHDITTTLDVWELKQVEEDKIEKVSDLKOLEEDILKEVKEI	1336
QY	609	----LIP--ASLXTI-----REAEFKKHLKQL--EVASALSHIAFNALDDNDGDQFDN	655
Db	1337	LSSEILEDYKELKTITDILKEKKEIEKHDPKEFEAEBEIKDLEADILKEVSSLEVBEE	1396
QY	656	KVVVKTHNSYALADGEHFI-----VDPDKLSSTIVLEKILKIEGLDYS	701
Db	1397	KLEEVHE--LKEEVEHIIISGDAHIKGLBEEDDLEEVDDLKSGILMDKGMELGDMDKE	1453
QY	702	TLRQTTQTOFRDWTTAGKALLSKSNLRQCEKQFLQEAOFFLGRVLDLRAIAKAE-----	756

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:40:26 ; Search time 45 seconds
(without alignments)
1672.139 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 5080

Sequence: 1 MKKHLKTVALTITVSVVTH.....LGYTSVALLSLTAIKKKY 1008

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 110 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*

2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*

3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*

4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*

5: /cgn2_6/ptodata/1/1aa/PTUS-COMB.pep:*

6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	215.5	4.2	641	4	US-09-583-110-4263
2	215.5	4.2	641	4	US-09-107-433-2948
3	214.5	4.2	3259	4	US-09-949-016-6507
4	213	4.2	1799	4	US-09-134-000C-5178
5	205	4.0	5024	4	US-09-710-279-2964
6	205	4.0	10182	3	US-09-134-001C-3159
7	204.5	4.0	324	4	US-09-248-796A-18798
8	199.5	3.9	2285	3	US-09-308-375-2
9	199.5	3.9	2285	4	US-09-932-183A-2
10	196.5	3.9	1588	5	PCT-US93-07261-11
11	196.5	3.9	1663	5	PCT-US93-07261-16
12	195	3.8	3696	3	US-09-134-001C-5080
13	192.5	3.8	2825	1	US-08-328-254-6
14	191.5	3.8	3248	1	US-08-353-700-1
15	191.5	3.8	3248	5	PCT-US95-16216-1
16	187.5	3.7	3210	4	US-09-538-092-1154
17	185	3.6	3878	4	US-09-914-259-11
18	183	3.6	1208	4	US-09-134-000C-5756
19	181.5	3.6	1786	3	US-08-973-462-8
20	179.5	3.5	928	4	US-09-134-000C-6590
21	178	3.5	1224	4	US-09-107-433-4347
22	177.5	3.5	1427	4	US-09-538-092-1044
23	176	3.5	1430	3	US-09-008-172-2
24	176	3.5	1430	3	US-09-210-361-6
25	176	3.5	1430	4	US-09-740-274-6
26	175.5	3.5	1078	4	US-09-248-796A-20284
27	175.5	3.5	1216	4	US-09-583-110-3824
28	173.5	3.4	173.5	4	US-09-248-796A-18993
29	173.5	3.4	2954	4	US-09-150-867-1
30	172	3.4	905	4	US-09-248-796A-16333
31	171	3.4	3241	4	US-09-841-786-1
32	169.5	3.3	1010	3	US-09-134-001C-5178
33	169.5	3.3	1965	4	US-09-583-110-3829
34	169.5	3.3	1972	4	US-09-107-433-3251
35	169	3.3	1095	4	US-09-107-532A-3855
36	169	3.3	1874	4	US-09-331-403-2
37	168.5	3.3	1027	3	US-08-446-137B-2
38	168	3.3	2343	3	US-09-324-867-2
39	167.5	3.3	584	4	US-09-949-016-10752
40	167.5	3.3	859	1	US-08-053-614-2
41	167.5	3.3	859	2	US-08-316-397B-2
42	167.5	3.3	859	2	US-09-034-306-2
43	167.5	3.3	859	3	US-09-259-437-2
44	167.5	3.3	859	5	PCT-US93-09782-2
45	167.5	3.3	1181	1	US-08-053-614-4
46	167.5	3.3	1181	1	US-08-316-397B-4
47	167.5	3.3	1181	2	US-09-034-306-4
48	167.5	3.3	1181	3	US-09-259-437-4
49	167.5	3.3	1181	5	PCT-US93-09782-4
50	167	3.3	872	3	US-08-851-843A-8
51	167	3.3	872	3	US-08-851-843A-54
52	167	3.3	872	3	US-08-974-549A-221
53	167	3.3	872	3	US-08-854-050-8
54	167	3.3	872	3	US-08-854-050-54
55	167	3.3	872	3	US-09-430-323-8
56	167	3.3	872	3	US-09-430-323-54
57	167	3.3	872	4	US-09-402-181B-221
58	167	3.3	872	4	US-09-721-456-221
59	167	3.3	872	4	US-09-766-253-8
60	167	3.3	872	4	US-09-766-253-54
61	167	3.3	2733	4	US-09-949-016-11433
62	166.5	3.3	1078	4	US-09-583-110-4036
63	166.5	3.3	1080	4	US-09-107-433-4843
64	166	3.3	2366	1	US-08-480-604A-10
65	166	3.3	2366	2	US-08-405-496A-10
66	166	3.3	2366	3	US-08-915-136-10
67	166	3.3	2366	3	US-08-957-310-10
68	166	3.3	2366	4	US-10-011-366-10
69	166	3.3	2366	4	US-09-084-517-10
70	165.5	3.3	862	4	US-09-248-796A-16013
71	164.5	3.2	1979	4	US-09-949-016-6468
72	164.5	3.2	2047	4	US-09-949-016-7404
73	164	3.2	1279	4	US-09-710-279-3188
74	163.5	3.2	1534	4	US-09-543-681A-5182
75	163.5	3.2	2411	3	US-09-268-347-36
76	162.5	3.2	1173	4	US-09-248-796A-19313
77	162	3.2	1964	2	US-08-790-912-3
78	162	3.2	2052	2	US-08-790-912-2
79	161.5	3.2	1282	4	US-09-134-000C-5785
80	161	3.2	886	4	US-08-956-171E-5235
81	161	3.2	886	4	US-08-781-986A-5235
82	161	3.2	990	2	US-08-645-193B-15
83	161	3.2	1132	4	US-09-248-796A-15026
84	161	3.2	1307	4	US-09-949-016-7561
85	160.5	3.2	582	4	US-09-081-149-8
86	160.5	3.2	956	4	US-09-107-532A-5007
87	160.5	3.2	1805	1	US-07-853-913-2
88	160	3.1	1493	4	US-09-713-273A-20
89	159	3.1	935	4	US-09-914-259-25
90	159	3.1	1164	4	US-09-538-092-399
91	159	3.1	1963	4	US-09-583-110-5243
92	159	3.1	1967	4	US-09-107-433-4883
93	158.5	3.1	1055	4	US-09-949-016-9776
94	158.5	3.1	1912	1	US-08-409-995-4
95	158.5	3.1	1912	3	US-08-685-467-4
96	158.5	3.1	2353	3	US-09-377-155-33
97	158.5	3.1	2353	3	US-08-913-942-4
98	158.5	3.1	2353	3	US-09-669-974-33
99	158.5	3.1	2353	4	US-09-797-862-33
100	158.5	3.1	2353	4	US-09-684-707-4

Sequence 18993, A
Sequence 1, Appli
Sequence 16333, A
Sequence 1, Appli
Sequence 178, Ap
Sequence 3829, Ap
Sequence 3251, Ap
Sequence 3855, Ap
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 10752, A
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 8, Appli
Sequence 8, Appli
Sequence 54, Appli
Sequence 54, Appli
Sequence 8, Appli
Sequence 54, Appli
Sequence 221, App
Sequence 221, App
Sequence 8, Appli
Sequence 54, Appli
Sequence 11433, A
Sequence 4036, Ap
Sequence 4843, Ap
Sequence 10, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 16013, A
Sequence 6468, Ap
Sequence 7404, Ap
Sequence 3188, Ap
Sequence 5182, Ap
Sequence 3, Appli
Sequence 2, Appli
Sequence 5785, Ap
Sequence 5235, Ap
Sequence 5235, Ap
Sequence 15, Appl
Sequence 15026, A
Sequence 7561, Ap
Sequence 8, Appli
Sequence 5007, Ap
Sequence 2, Appli
Sequence 20, Appli
Sequence 25, Appli
Sequence 399, App
Sequence 5243, Ap
Sequence 4883, Ap
Sequence 9776, Ap
Sequence 4, Appli
Sequence 33, Appli
Sequence 4, Appli
Sequence 33, Appli
Sequence 33, Appli
Sequence 33, Appli

101 158.5 3.1 2354 3 US-09-268-347-47 Sequence 47, Appl
102 158.5 3.1 3289 2 US-08-477-451-2 Sequence 2, Appl
103 158 3.1 1074 4 US-09-071-035-358 Sequence 358, App
104 158 3.1 1074 4 US-09-071-035-394 Sequence 394, App
105 158 3.1 1972 4 US-08-875-435B-3 Sequence 3, Appl
106 158 3.1 2663 4 US-09-538-092-1252 Sequence 1252, Ap
107 157.5 3.1 1001 4 US-09-248-786A-18658 Sequence 18658, A
108 157.5 3.1 1164 3 US-08-923-992A-2 Sequence 2, Appl
109 157 3.1 990 2 US-08-392-625-20 Sequence 20, Appl
110 157 3.1 990 2 US-08-466-961A-20 Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-09-583-110-4263
; Sequence 4263, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4263
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4263

Query Match 4.2%; Score 215.5; DB 4; Length 641;
Best Local Similarity 23.2%; Pred. No. 8.7e-07;
Matches 132; Conservative 72; Mismatches 178; Indels 187; Gaps 30;

QY 6 KTVALTTLTVSVV-----THNQVFS-----LVKEPILQK 35
DB 3 KTTLSLTTAAVILAAVYVNEPILADTPSEVVKETKVGSIIOQNNIKYKVLTVEGNIRT 62
QY 36 TOASSISGADY-AESSGK-----SKLXINETSGVDDTVTDLFS-----DK 76
DB 63 VQVGVNGVTPVEPAGQDGKPTIPTKITVGDKVFTVEVASQAFSYYPDETGRIVYYPSS 122
QY 77 RTTPKIKDNLAKGPREQELKAVTENTSEKQITSGSQLEQSK-----120
DB 123 ITIPSSIKKIQKGFHSGKAKTII-----FDKGSQLEKIEDRAFDFSELEBELPAS 174
QY 121 -----SLSLANK-TVPSTSNWEICDFITKGTNLVGLSKGVKLSQTDHVLPSQ 168
DB 175 LEYIGTSAPFSQKSLKLTFFSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222
QY 169 AAD-GTQLIQVA-----SPA-----FTPDKKTAIAEYTSRAGENGESQLD 208
DB 223 VKTLSGNLFRLTSLKHVDVEEGNESFASVDGVLFSKQ-KTQLIYPSQ--KNDESYKTP 279
QY 209 VDGKEIINEGEVFNYSLLKKVTIPTGYKHIGDQAFVDNKNIAEVLNPESLETISYAF-A 267
DB 280 KETKELASVSFNKNSY-LKKLELNEGLEKIGTGFADAKLEIEISLPNSLETIERLAFVG 338
QY 268 HIALKQIDLPDLNKAIGELAF-----FDNQITGKLSLPQRLMLAERAFKSNHIK 317
DB 339 NLELKELILPDNVKRVFGHVMNGLPKLKSALTGGNNIN---SLPSPFLSGVLDSLKEIHK 395
QY 318 -----TIEPRGNS--LKVIGEASFQ-DNDLSQLMLPDGLEKIESAFT 357

Db 396 NKSTFSPKKDTFALPETVVKFYVTSEHKVQLKSNLSTSNDI-----IVEKVDNIKQETDV 451
QY 358 GNPDDHYNRRVWLWTK-----SGKNPSSGLATENTYVNPDKSLM---QSPSPIDY 404
Db 452 AKP-KKSNQGVGVWVKDKGLWYLYLNESSGSMATGMV-----KDKGLWYLYLNESSGM-A 502
QY 405 TKWLEED--FTYQKNS-----VTGP-SNKGKL 427
Db 503 TGMVKDKGLWYLYLNESSGSMATGMVKDKGL 531

RESULT 2
US-09-107-433-2948
; Sequence 2948, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2948:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 641 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...641
; SEQUENCE DESCRIPTION: SEQ ID NO: 2948:
US-09-107-433-2948

Query Match 4.2%; Score 215.5; DB 4; Length 641;
Best Local Similarity 23.2%; Pred. No. 8.7e-07;
Matches 132; Conservative 72; Mismatches 178; Indels 187; Gaps 30;

QY 6 KTVALTTLTVSVV-----THNQVFS-----LVKEPILQK 35
DB 3 KTTLSLTTAAVILAAVYVNEPILADTPSEVVKETKVGSIIOQNNIKYKVLTVEGNIRT 62
QY 36 TOASSISGADY-AESSGK-----SKLXINETSGVDDTVTDLFS-----DK 76

Db 63 VQVNGVTPVEFAGQDGKPFPTIPTKITVGDKVFTVTEVASQAFSPYDPTGRIVVYPS 122
QY 77 RTTPEKIKONLAKPREQELKAVTENTSEKQITSGSOLEQSK----- 120
Db 123 ITIPSSIKIKQKGFSGAKTII-----FDKGSQLEKIEDRAFDFSELEELPAS 174
QY 121 -----SLSLNK-TVPSTSNWEICDFTKGNLTVGLSKSGVEKLSQTDHLVLP 168
Db 175 LEYIGTSAPFSQKAKLTFSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222
QY 169 AAD-GTQLQVA-----SFA-----FTPDKKTAEVTSRAGENGESOLD 208
Db 223 VTKLSNLFRLTSLKHVDVEGNEFASVDGLFSKD-KTQLIYYPQ--KNDESYKTP 279
QY 209 VDCKEIIINGEVFNYSLLKKVTIPTGKHIGODAFVNDKNIAEVLNPSLETISDYAF-A 267
Db 280 KETKELASVFNKNSY-LKLELNEGLEKIGTFAPADAKLBEISLPNSLETIERLAFYG 338
QY 268 HIALKQIDLPDNLKAIGELAF-----FDNQITGKLSLPRQLMRLAERAFKSNHIK 317
Db 339 NLELKEILLPDVNFKNFGKHMNGLPKLSLTIGNNIN--SLPSPFLSGVLDLSLKEIHK 395
QY 318 -----TIEFRGNS--LKVIGEASFQ-DNDLSQMLPDGLEKIESEAPT 357
Db 396 NKSTEPSVKOTFAIPETVKFYVTSHEIKDVLKSNLSTNDI-----IVEKVDNIKOETDV 451
QY 358 GNPGDHYNRVVLTWK-----SGKNPSGLATENTYVNPDKSLW---QESPEIDY 404
Db 452 AKP-KKNSQGVGVKDKGLWYLNESGSMATGW-----KDKGLWYLNESGSM-A 502
QY 405 TKWLEED--FTYQKNS--VTGF-SNKGL 427
Db 503 TGVVKDKGLWYLNESGSMATGWVKDKGL 531

RESULT 3

US-09-949-016-6507
; Sequence 6507, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6507
; LENGTH: 3259
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6507

Query Match 4.2%; Score 214.5; DB 4; Length 3259;
Best Local Similarity 19.8%; Pred. No. 1.1e-05;
Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;

QY 20 HNEVFSLVKPEILKQTOASSISGADYAESGSKSLKI-----NETSGPVDPTVDL 72
Db 233 HEDELLQLV-----TQAD-----VETENQKRLVLRQKLEHESLSVGRAGVVDL 277
QY 73 FSDKRTTPEKIDNLAGPREQLK-----AVTENTESEKQITSGSOLEQSKESLSLNK 126
Db 278 LOBELTAABORNOILSQQLQMEAEHNTLRNVTETERESEKILLERKMEVAERKLSF-- 336
QY 127 TVPSTSNWEICDFTKGNLTVGL-----SKSGVEKLSQTDHLVLPQAAADGTGLIQV 178

RESULT 4

US-09-134-000C-5178
; Sequence 5178, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032

Db 337 NLQEMHLLLEQFEQAGQAQAELESRYSALEQKHKAEMEKTSHLSLQKTG---QELQS 393
QY 179 ASFPATPPDKTAIAEYTSRAGENGEL-----SOLDVDGKEIINEGEVFNYSLLKKVTIPT 233
Db 394 ADCALDKQNSKLLQDNEQAVQSAQTIQOOLEDQLOQSKEL-----SQFLNRLPQQ 445
QY 234 GYKHIGQDAFVNDKNIAEVLNPSLETISDYAFALHALKQIDLPDNLKAIGELAFDNOI 293
Db 446 -HETASQTSFPDVYN-----EGTQAVTEENIASLQKRVVL-----ENEK 484
QY 294 TKGLSLPQLMRL-AERAFKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKTE 352
Db 485 GALLLSLSELEBELKAENKLSQITLLEAQNR-----GEA-----DREVSEISIVDIANKS 537
QY 353 SEAFTCNPGDHHYNNRVVLTWKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDF 412
Db 538 SSA-----BESGD-----VLETFQSOKHEL-----SVLLEMEKAQEERI 573
QY 413 TYQKNSVTG-----FSNKGLOKVKRNKNLE-----IPKQHNVT 446
Db 574 AFLKLQLOKRAEADHEVLDQKEMKQMEGEGIAPIKMKVFLDTGQDPLMPNBSLIP 633
QY 447 ITEIGDNAPRVNDPQNTLRKYDLSE--VKLPSTIRKIGAFQSNLNKLSFEASDDLEI 504
Db 634 AVE--KEQASTEHQSRTESEISLNDAGVELKST-----KQDGDKLSAVPDIGQC 681
QY 505 KEGAFMNRRIETLEIKDLKLVITIGDAAFH-INHIYAIVLPESVQEIQR-----SAFRQCA 558
Db 682 HQDELERLKSQILEL-----ELNFKHAQEIYEKNDLDEKAKEISLNLQLEEFKKNAD 733
QY 559 NN-----LIFWGSKVTKLGEMAFNSNLEHLDL-----SEOKLTEIPVQAFS 601
Db 734 NNSAFTALSEERDQLL---SOVKELSMVTELRQVKQLEMLAEARORRLDYESQTAH 790
QY 602 DNALKEVL---LPASLKTIREEAFKGNHLKOLEVASALSHTAFNALDDNDGDQPDNK- 656
Db 791 DNLLTEQIHSLSIEAKSKDVKIEVL-QNELDDVQLQFSEOSTLIRSL-----QSOLONKE 844
QY 657 -----VVVTHNSYALADGEHFIVDPKLSSTIVDLKILKLEGLDYSTLR 704
Db 845 SEVLEGAEVRHRISSKVEELSQALESQLEITKMDQLL-----LEK-----KRDVETLIQ 893
QY 705 OTTOFTQFDMTTAGKALLSKNLQGEKQKQFLOEQAFFLGRVLDLQAKAEKALVTKA 764
Db 894 QTIEEKDQOVT-----EISFSTERKQVQNEEFSLG-VEI-KTLKQNLNLSRAE 943
QY 765 TKNGQLLERSINKAVLAYNNSAIKKA---NVKRLKEKELDLTLGLVEGKPLAQATMVQGV 821
Db 944 AKKEQVEEDNEVSSGLKQNYDEMSPAQGISKEELQHEFDLLKKEQKRLQKLAALINRK 1003
QY 822 YLLKTPPLPEYIIGLVNVPKSGKLIYALDMSDT-IGEGQKDAYGNPILNVDENEGVH 880
Db 1004 ELLQVRSRLEBELANLK---DESKK---EIPLSETERGEVEE-----DKNKEYS 1047
QY 881 ALAVATLADYEGLDIKTILNSKLSQTSIR-----QVPTAAVHRAGI 922
Db 1048 EKVTSKQCEIYILKQTISEKEVELQHIRKDLKLEKLAESQFQALVKQMNQTLQDKNQ 1107
QY 923 FOAIQNAAAAEQQLLPKFTHSEKSSSES-ANSKDRGLQSNPKTNRGH 971
Db 1108 IDLLQAEISENQAIQKLTSTNTDASDGDGVALVKETVVISPPCTGSSEH 1157

```

; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5178
; LENGTH: 1799
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5178

Query Match
Best Local Similarity 4.2%; Score 213; DB 4; Length 1799;
Matches 239; Conservative 176; Mismatches 423; Indels 398; Gaps 56;

QY 49 ESSGSKLKINETSIPVDDTVTDL-----PSDKRTT-----PEKT---83
D 684 EAQGAATTKIN--SQGPKNGEKDGLTVTTEPTKSGYKDFVDERSTRDLDLASEAPESVTAM 742
QY 84 ----KDLAKGREBELKAVTENTSEKQI-----TSGSOLEQKESLSLKNKTPVST 131
D 743 VGGTEVTLQAANKEEFLAGTNLYYFDKFPQVQNVLSASEKLNQSAKQVIA 802
QY 132 SNWEIC--DFITK-----GNTLVGLSKSGVEKLSQTDHLVLPQAA-----170
D 803 KDVQITVKGFINKGTVDGGNTTV-----DDQLTIPANVAINEEKTTPSSLTLQ 850
QY 171 -----DGTQLIQVASFAPDPKKTATAEYTSR---AGENG--EISQ-----206
D 851 WQVTEATSYEVRDGTVPNGINTATFDGFSPLSEHTFPRVAVGKNGVSESEPIKKG 910
QY 207 -LDVDSKEIINGEVFNSY-----LLKKVT---IPTGYKHIGQDAFVDNKVIAENVLPE 256
D 911 TQDDPYKETIINGKATSNLPEQPGALKKLTDKDLSTGW-----HTWST 955
QY 257 SLETISDYAFALHALK-----QID-----LPDNLKAIGEL-----AFFDN 291
D 956 GIANPSDGNF--LSLKFDLGAEYQMDKIEYLPDRNAGNGNLTQYRTSKDGANWTFSE 1013
QY 292 QITGKLSLPRQLMRILAEAFKSNHKTTFEFGN-----SLUKVTEA 332
D 1014 PINWKQDALTKTIETKQAYRVEVMKVLKSVGNFGSGREMLFYKPGTGEILHGDITNDG 1073
QY 333 SFQDNDLSQLMPLDGLKIESEAFG--NPGD-----DHVNRVVLWTKSG-----376
D 1074 TIDENDAMSYRNYTGLESVDS--FNGYVEKGLNKGVIDAYDISYVLRLQDGGIEIPDV 1132
QY 377 -----KNPSGLATENTYVNPDKSLWQSPE 401
D 1133 ERIAGLSLAVNENGKOTYLPDGLTFLTKQDGLKNALSTKMSFSSKFLVGQPAT 1192
QY 402 IDYTKWLEBDFYQKNS-----VTGFSNKGLOKVRKNLEIPKQNGVTTIEGDNF 455
D 1193 TNNTOOMENYSYKRXHSDVENLYLVLSNOG-----NKQL-----LNG-----1230
QY 456 RNVDFONKTLRKYDLEEVKLPTIRKIGAF-----AFQSNLKSFEASDDLBEIK 505
D 1231 -SMDLVFVKVKTETRVKRTVTEQPIQFQMSQGLLVGQGFQQAATLSDFSVT-----1282
QY 506 EGAFMNRITELKDK-----LVTIGDAAPHIN-----IYAIVLPESVQEIERSAFR 554
D 1283 -----VKPTFLVDVDELQALITLNQARVEKYTPETWALFKPILDEAVAVLANEQAT 1334
QY 555 QN-----GANNLIFMGSKVKTLGEMAFNLRLHLDLSEKQLTPIPV--QAFSDNALK--606
D 1335 QTDVSAAEENLEKAAASOLEKMPDVA---NKAD-LEKAIQELAKKPSDQGEFTETTKVL 1390
QY 607 EVLLPASTLTIRBEAFKQNHKLQLEVASALSIAFNALDDNDGDEQDPNKVVVKTTHNSY 666
D 1391 EESLAAQKVFQAEKVTOBEIDQ-----ATKTLREAI 1422
QY 667 ALADGHEHFDVDPDKLSSTTVLEKILKLEIGLDYSTLROTTOTOFDMTMTAGKALLSKSN 726

```

Db 187 ILTKNALDGEOLRAKENADQEIINTLNQTDQRNSEKGLVNSQTRTEVASQALAKX 246
Qy 213 EIINEGEVNSYLLKKVTIPTGYKHICQDAFVQNKNAEVLNIPESLETISDVAFAPALX 272
Db 247 ELNKVMEQLNLLNGKNQMINSSKFINEA---NQOAYSNATASAEVLKXKS---QNP 299
Qy 273 QIDLPNLKAIGELAFDQITQKLSPLQMLAERAFKSNHKTIEFRGNSLKVIGEA 332
Db 300 ELDKVTIEQAINNINSAIINLNGEAKLTKA-----KEDAVASI----- 337
Qy 333 SFQDNLSQLMLPDGLEKIESEAFNPGDDHYNVRLVTKGKPNKSLATENTVNP 392
Db 338 ----NNLSC---LTNEQTKENQAVGSTRDQVAN---VLRDSKALDQS---MOTLRDLVNQ 388
Qy 393 KSLWQSSPEIDYTKMLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPKQHGVTIIEGD 452
Db 389 NVIHSFS-----NYFNEDST-QKNTYDNDALNGSTYITCOHNSLNKSTIDTISQI-- 439
Qy 453 NAFRNVDPQNTLURKYDLEEVKLPSTRIGAFAPAFOSNNLKGFAS-----DDLEB- 503
Db 440 NTAKN-DLHGAELQDRDKGTAN-----QBITGQLGYLNDPQKSAEESLVNGSNTRSEVEH 493
Qy 504 IKGAFMNRRIETLEKDKLVT-----IGDAAFINHIYAIVIPES-----VQIEGR 550
Db 494 LNEAKSLNAMK---QLRDKVAEKTNYKQSSDYINDSTEH---QRGYDQALQEAENIINEIGN 550
Qy 551 SAFRQNGANNLIFMGSKVTKLGMFLSNRLE---HLDLSEQO---LTEI-PVOAFSD--- 602
Db 551 PTLNKSEIE-----QKLOLTDQNALQGSHL-LBEAKNNAITEINKLTALNDAQR 600
Qy 603 ----NALKEVLLPA-----SLKTIREAFKKNHLKOLEVASALSHTAFNAL 644
Db 601 QKAIENVQAQOTIPAVNQOLTLDRINTAMQALRDKVGGQNNVHQ-----QSNYFN-- 651
Qy 645 DONDGEQPDNKVVVTHNSVALADGEHFIVDPDKLSSTIVDLEKILKIEGLDYSLR 704
Db 652 EDEQPKHNDNSVQA-----GOTII---DKLODFIMNKNEIEQAIIQIN----- 692
Qy 705 QTTOTQFRDMWTAGKALLSKNSRQEGOKFLQEAQFFLGRVDLDKAIKAEKALVTKA 764
Db 693 -TTOT-----ALSGENKLTDOESTNRQ-----IEGLSLNTAQINAEKDLVNOAK 737
Qy 765 TKNQOLLERSINKAVLAYNNSAIKKANVRLEKELDLLTGLVEGKPLAQATWVQGVYLL 824
Db 738 TR-----TDVAQKLATAKEINSAMSNLRDGIQNKEDIKRSS-----AYIN 777
Qy 825 KTPPLPEYYIGL-----NVYFDKSGKLIYALDMSDTIGEGQDKAYGNPILVD 873
Db 778 ADPTKVTAYDQALQNAENIINATPNVELNKA-----TIEQAL-----SRVQQAQ 821
Qy 874 EDNEGVALAVA-----TLADYEGLDIKTILNKSLSOLTSIRQVPTAAVHRAG 921
Db 822 QDLDGVQOLANAKQATQTVNGLSLNDQCKRELNLLINSANTR-TKVOEELNKATESNH 880
Qy 922 IFQAIQAAAEABQLL-----PKPGTHSEKSSSES 952
Db 881 AMBALNSVQNVQVQKSNYNNEDQPEQHNDYNAVNEA 919

RESULT 6
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 4.0%; Score 205; DB 3; Length 10182;
Best Local Similarity 19.1%; Pred. No. 0.00029;

Matches 201; Conservative 178; Mismatches 447; Indels 224; Gaps 45;
Qy 47 YAESGKGLKINETSGPV--DDTVDFSDKRTTPEKIKONLAKGPRQEQLKAVTENTE 104
Db 7618 YNQAIVNAKAKNIINDQPTPMANDEIQSVLNEVKQT---KQNL-----HGDQKLANDKTD 7668
Qy 105 SEKQITSSQOLSESLNKTIVSTSNWEICDFITKGTNLVGLSKSVGEKLSQTDHLV 164
Db 7669 AQAATLNALNYLNAQQRGNLETKVQNSNSRPEVQKVVLQANQANDAMKCLDDALTGNDAIK 7728
Qy 165 LPSQAADGTQLIQVASFAPTPDKTAIAEYTSRAGENGESOL-----DVGKELI 215
Db 7729 QTSNVINEDTSQVNFDEYTRGKNIVAEQTPNPNMPTNINTIADKITEAKNDLHGVLK 7788
Qy 216 NEGEVFNYSLLKKVT-IPTGYKHIGQDAFVDNKNIAEVN-----LPESLETISDYAF 266
Db 7789 KQAQOQSINTINQMTGLNAQKQEQLEQIBEQTOTRSEVHOVINKAQAALNDSMTLRQSIT 7848
Qy 267 AHLAKQDLDLNLKAIGELAFPDN-----QITKLSLPLQMLAERAFKSNHKTIE 320
Db 7849 DEHEVYKQTSNYIN-ETVGNQTAYNNAVDKVGQIINQTSNPTMNPLEVERA-----TSNVK 7902
Qy 321 PRGNSLKVIGEASFQDNDLSQLMLP---DGLKIESEAFGTG-----NPGDDHYNVNRV 369
Db 7903 ISKDALH--GERELNDNKNKTFVAVNHLNLMQAOKEALTHEIQATIVSQVNNIYNKAK 7960
Qy 370 VLMTKSGK-----NPSGLATENTYVNPD---KSLWQE-----SPEIDYTKW 407
Db 7961 ALNDNMKKLKDIVAQDQNVRSNNYINEDSTPQNNYDINHQAQSIIDQVANPTASHDE- 8019
Qy 408 LEEDFTYQKNSVTGFSNK-GLOKVRKNKLEIPKQHGVTITEIGDNAFRNVDFQNTLR 466
Db 8020 IENAINNKHAINALDGEHLKQQAENANL-----LINSNLNLAQPDARNLVN 8070
Qy 467 KYDLEEVKLPSTRIGAFAPAFQSNLKSFEASDDELEIEKEGAFMNRRIETLEKDKLVTI 526
Db 8071 EAQTRF-----KV-----AEQLOQAQALNDAMKHLNRNSIQNQ---SSVQESKYINA 8114
Qy 527 GDA-AFHINHIYAI-----VLPESVQIEGRSAFRQ-----NGANNLIFMGSKVKTILGEMA 575
Db 8115 SDAKKEQYNH-AVREVENIINEQHPDLDKEIKQLTDGVNQANDL----- 8159
Qy 576 FLNSNLEHLDSQEQQLTEIP---VQAFSONALKE-----VLLPAS 613
Db 8160 ---NGVELLDADQNAHQSIPTLMLNQAOQNALNEKINNNAVTRTEVAALIIQOAKLLDHA 8216
Qy 614 LKTIEEAFKKNHLKOLEVASALSHTAFNAL-DNDGDGEQPDN-----KVVKVTHNSY 666
Db 8217 MENLEESIKDKEQVKQ-----SSNYNEDSDVOETVDYDNAVHVTILNQTVNPTL 8266
Qy 667 ALADGEHFIVDPDKLSSTIVDLEKILKIEGLD--YSTLRQTTQTFQFRDMTTAGKALLSK 724
Db 8267 SIEDIEHAINEVNOAKQLRGKQKLYQITDLADKELSKLDDLTSSQSSISINQIYAKTR 8326
Qy 725 SNLRGQ-BKQPLQBAQFPLGRVLDLDAIKAAKALVTKKATKNGQLLERSINKAVLAYN 783
Db 8327 TEVAQAIKAKSLNHAM-----KALNKVYKNADKVLDSRFINEDQPEKAYQQAINHVD 8381
Qy 784 NSAIKKAN-----VKLEKELDLLTGLVSGKGLPAQ-----ATWVGVLKTPPLP 831
Db 8382 STIHRQTPNEMDPTVINSTIHELETAQNHLHGDKLAAHQDAAANVING--LIHLNVQAR 8439
Qy 832 EYIIGLVNVFQKSGKLIYALDMSDTIGEG-----QKDAYGNPILN-----VDEEDNGYHAL 882

```
Db      8440  EWMINTNTATREKAVKLDNAQALDKAMETLQVVAHKNNILNDSKYLNEDSK-YQQQ 8498
Qy      883  AVATLADYEGDLIKTILNSKLSQTSIRQVPTAAVYHRAGIFQAIQNAABAEQQLPKPGT 942
Db      8499  YDRVIADAEQL-----LNQTTN-----PTLEPKYVDIVK--DNVLANEKILP----- 8538
Qy      943  HSEKSSSESANSKDRGLQSNPKTNRGHS 972
Db      8539  GAELKSYDKS-NANDEIKHMYLNNAQKQS 8567

RESULT 7
US-09-248-796A-18798
; Sequence 18798, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18798
; LENGTH: 924
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18798

Query Match      4.0%; Score 204.5; DB 4; Length 924;
Best Local Similarity 20.2%; Pred. No. 8.9e-06;
Matches 184, Conservative 163; Mismatches 351; Indels 213; Gaps 46;
Qy      178  VASFAPTDPKKTAI---ABYTSA---GENGEISQLDVDGKEIINEGEVFNLYLK--KV 229
Db      20  VVTFFYQNTRTTFVDMATTTTSNAPANGENASITQIIVNHKNVSKADDLISFNKQKL 79
Qy      230  TIPTGYKHIGQDAFVDNKNIAEVLNPELSTSDYAFAPHALKQIDLPNLKAIGELAFF 289
Db      80  KVVNLDPYDGSFAFIDS-----SUKVSDP-----VKIPEPAN-SPVNSLAKT 122
Qy      290  DNQITGKLSPLQMLRL-----AERAFKSNHIKTIFFRGNS---LKVTGEASFQDN 337
Db      123  NND-----TVTKQVRVPSYTRPTLSSSRHSPKNLSKRSRINSNSPDKMTLSKRVRLKNS 177
Qy      338  DLSQLMLPDGLEKISSEAPFTGNPDHNNRVVLTWKSGKNPSGLATENTYVNPDP----- 392
Db      178  PLSQ-----SVDSNPPI-----EKNSKIYL-----KN-LALRTSSTDTLTDSGSH 217
Qy      393  -----KSLMQB-----SPEIDYTKWLEDFYQKNSVTGFNS-----KGLQVKRNKN 435
Db      218  DCCLQIETLQNEKRVHLHKYVDSIKDLQDFLAFQKQEAAGIINVLDLDAQLSELKQ-KQ 276
Qy      436  LEIPKQHNQVTTIEIGDNAFRNVDFQNTLRKYDLEEVKLPSTIRKIGAFQSNLKSF 495
Db      277  IDLENTKNDQIKELVSEHLOLEKISKOLTEKFKLVETQLLSK-HESHVQQTKELIAB 335
Qy      496  RASDLEETKEGAFMNNRIETLELKKDL-----VTIGDAAPHINHIIVLPSVQSIG 549
Db      336  ESKQVQVEELE-----NLKTHAKDSARILELETQLSDAA-----KEKS 374
Qy      550  RAFPQNGANNLI-FMGSVKVTLGEMAFLSNLEH-----LD-LSEQKOLTIPIQA 599
Db      375  ESDYKLTDTSEIVNDLKSQIETL--KANL-NKLEBEREIQNKGLQVSELKELKLVEE 431
Qy      600  FSDNAL-----KEVLLPASLKTIRBEAFKKNHLKQLEVASALSHTAFNALDND--- 648
Db      432  LSNLLKQLQELHRKEIELNEQLKLGVSADKELKLSQISQ--SKIYINSLKQNAKT 489
```

```
Qy      649  -----GDEQPDNKVVVTKTHNSYALADGEHEFIVDPDKLSSTIVDLE---KILK--LI 695
Db      490  DRNWEIKYGTLEQNOQKLVBEH-----LANRQDQLEIKLVKSKTELEADNQAQKNQLT 544
Qy      696  E-----GLDYETL-RQTTOTQFR-----DMWTAGKALLSKNSLRQGEKQKPLQ 737
Db      545  EISKQFSTLGVSESELKKQLEBETQIKYDQAEIVHQELKAQIEKLLKESAGKQGLKELNQ 604
Qy      738  EAQFFLGRV--DLDKAJAKAEKALVTKATKNGQL-----LERSINKAVLAYNNSAIKXAN- 791
Db      605  SHEKFVNELKNEHEKQVKETKDQIIKEMBEKHQQAIEKIEDSHSHENIKKEINNEHENAKC 664
Qy      792  -VKLELEKELDLTLGLVEGKGPLAQATWVGVLTKPLPLPEYVIGLVNVPYDKSGKLIYA 850
Db      665  IISLNEIEBELTSQL--KNAESEKNTLQSLK-----EYENETIAYKSKIDQL--- 711
Qy      851  LDMSDTTIGEGOKDAYGNPILNVNDENGYHALAVATLADYEGDLIKTILNSKLSQTSIR 910
Db      712  ---EKESAENLKE-YEAKLOSMKFDLESDLAIEKQLPKD-DGQDFENQI-EKLNQLVTEK 765
Qy      911  QVPTAAVYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGR 970
Db      766  DLQLEKSD-----EEIASIKKMEBELDTKAKLEKSTTQNGMKLLDNLSTK-----K 813
Qy      971  HSAILPRTGSK 981
Db      814  HAAVVAELSGK 824

RESULT 8
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match      3.9%; Score 199.5; DB 3; Length 2285;
Best Local Similarity 20.9%; Pred. No. 7.7e-05;
Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;
Qy      81  EKIKNLAKGPREQELKAVTENTES-EKQITSGSQLEQSKESLSLN----- 125
Db      771  KKAKDDPEQS--QQTNVEAITTKDSTDKLIQOYKELQKVKESSRLTSDEBEQYLVQVTOQL 829
Qy      126  -KTVPSTSNWEICDFTKNGTTLVGLSKSGVEKLSQT-DHLVLPQAA-----DGTQ 174
Db      830  AQTTPAL-----VKGYDSQGNAILTKNKELEKAIENKEYLALKKQETRDSAKKTFEDASK 885
Qy      175  LIQVASPAFTPDKTAIABYTSRAGENGETSQLDVDGKEIINEGEVFNLYLKVVTTPTG 234
Db      886  EIKKSKDELKQYKQ--IADYNDKGRPKWDLTADDDYKVAADKAK--QSMUKAQSDIESG 941
Qy      235  YKHIGQDAFVDNKNIAEVLNPELSET-ISDYAFAPHALKQIDLPNLKAIGELAFFDNOI 293
Db      942  NAKVDSVLSTANAYSSIDISNTLKTSISD-----VVKLNKLDLDDLP-BELEKFFSSSL 994
Qy      294  TGKLSLPRQLMRLAERAFKSNHIKTIFFRGNSLKVIGESAFQDNDLSQMLPDGLEKTES 353
Db      995  -GKLEKMQ-----KALDSGEKAFD---NAKK-----DLQSL----- 1024
```

```
QY 354 EAFTGNPGDDHNNRVVLMTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFT 413
Db 1025 -----ETYSKSDSS-----IDVPKMSFD--K 1043
QY 414 YQKSVTGSNKGLOKVKRNKNLEIPKQHNGVTITIEIGNAFNVDVFNQKTKRYDLEEV 473
Db 1044 AQKNIKDQ--DKSLSSVKSEVG-----DLGETLABAGNEA---EDFGKK-----LKEA 1086
QY 474 KLPSTIRKIGAPAFQSNLKSPEASDLEBEIKEGAFMNNRIETLEKDKLVITIGDAAFHI 533
Db 1087 LDANSVDDIKAAIKEMSDAMQF--DSVDVLNGDIFNN-----TKDQVAPLND----- 1132
QY 534 NHIYAIVLPESVOEIGRSAFRQNGANNLIPMGSKVKTGEMAFNLSRHLHDLSEQKLT 593
Db 1133 -----LLEKMAE--GKS--ISANEANTLI---QKDELAQAISIEGVVVKINRDEVIKQR 1180
QY 594 EIPVQAFSD-----NALK-----EVLIPASLKTIRE-----DYSTLR 704
Db 1181 KVKLDAYNDMVTYSNKLMTKEVNNAIKTLNADTLRIDSLKLRKERKLDMSAEALSDELV 1240
QY 621 -----AFKNHKLKOLE-----VASAL--SHIAFNALDDNDGDQF 654
Db 1241 KSINNVDADAKKELKLEKMLQPGGYSNSQIEAMQSVKSALESYI--SASEATSTQEMN 1298
QY 655 NKVVVTHNSYALADGEHFIIVDPDKLSSTIVD-----LEKILKLEI-----DYSTLR 704
Db 1299 KQALVEAGTSLNWTDDQEKANEETKTSMYVVDVKYKEALEKVNABIDKYNKVQNDVPKYS 1358
QY 705 QTTQTQFRD--MTTAGKALLSKNLRQGEKQKFLQEAQFPLGRVLDLKAIAKALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQOKKKLMQ--EQAKLLKD--QIKSGNI-----TQYGIWTS 1404
QY 764 ATKNG--QLLERSINKAVLAYNNSAIKKNVNRLEKDLTLGLVEGK--GPLAQATMVOG 820
Db 1405 TSSGGTSPSTGSGYSKYSYNSAASKYNV-----DPALIAAVIQOESGFNAKARSGVG 1459
QY 821 VYLLKTPLPPEYIYGLNVYFDKSGKLIYALDMSDTIGEGOK--DAYGNPILNVDEBNEY 879
Db 1460 AMGLQMLPATAKSLGVNNAYDP-----YQNVMGTKYLAQOLEKEGG--NVEKALAY 1511
QY 880 HALAVATLADYEGLD-----IKTIL---NSKLSQTSIROVPTAAYHRAGIFOAION 928
Db 1512 NA--GPGNVIKYGGIPPEKFTQNVVKIMANYSKLSATS-----SIASY----- 1556
QY 929 AAAEAQILPKPQTHSEKSSSESANSKRGLOSNPKNRGRSAILPRTGSKG--SFVY 986
Db 1557 -----TNSAFRVSSKYQOBSGLRSP--HKGTDFAAKAGTATKSLQSGKV 1601
QY 987 GILGYTSA 995
Db 1602 QIANGSKTA 1610

RESULT 9
US-09-932-183A-2 ;
; Sequence 2, Application US/09932183A
; Patent No. 6832265
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
```

```
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-932-183A-2

Query Match 3.9%; Score 199.5; DB 4; Length 2285;
Best Local Similarity 20.9%; Pred. No. 7,7e-05;
Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;

QY 81 EKIKNLAKPPEQELKAVNTES--EKQITSGSQLEQSKESLSLN----- 125
Db 771 KKAKDDFEOS--QQTNVEAITTKDSDTKLIQYKSLQKVKRSRSLTSDBEQEYLVQVTOQL 829
QY 126 -KTVPTSTSNWEITCDFTKTNTLVGLSKSGVEKLSQT--DHLVLPQAAA-----DGTQ 174
Db 830 AQTFPAL-----VKGYDSQGNALTKNKELEKAIENTKEYLALKKQTRDSAKTTFEDASK 885
QY 175 LIQVASFAPTPDKTAIAEYTSRAGENGESISOLDVDGKEIINEGEVFNYSLLKKVTIPTG 234
Db 886 EIKKSKDELKQYKQ--IADYNDKGRPKMDLIAADDDDKYVAADKAK--QSMLKAQSDIESG 941
QY 235 YKHIQDQAFVDNKNTAEVNLPESET--ISDYAFALHAKQIDLPNLKAIIGELAFEDNOI 293
Db 942 NAKVKDSVLSTIANAYSSIDISNTLTKTSISD-----VNVKLNKDDOLDP--EELEXPSSSL 994
QY 294 TGKLSLPRQLMLRAERAFKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIES 353
Db 995 -GKLEKMQ-----KALDSGDEKAFD---NAKK-----DLQSLI----- 1024
QY 354 EAFTGNPGDDHNNRVVLMTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFT 413
Db 1025 -----ETYSKSDSS-----IDVPKMSFD--K 1043
QY 414 YQKSVTGSNKGLOKVKRNKNLEIPKQHNGVTITIEIGNAFNVDVFNQKTKRYDLEEV 473
Db 1044 AQKNIKDQ--DKSLSSVKSEVG-----DLGETLABAGNEA---EDFGKK-----LKEA 1086
QY 474 KLPSTIRKIGAPAFQSNLKSPEASDLEBEIKEGAFMNNRIETLEKDKLVITIGDAAFHI 533
Db 1087 LDANSVDDIKAAIKEMSDAMQF--DSVDVLNGDIFNN-----TKDQVAPLND----- 1132
QY 534 NHIYAIVLPESVOEIGRSAFRQNGANNLIPMGSKVKTGEMAFNLSRHLHDLSEQKLT 593
Db 1133 -----LLEKMAE--GKS--ISANEANTLI---QKDELAQAISIEGVVVKINRDEVIKQR 1180
QY 594 EIPVQAFSD-----NALK-----EVLIPASLKTIRE----- 620
Db 1181 KVKLDAYNDMVTYSNKLMTKEVNNAIKTLNADTLRIDSLKLRKERKLDMSAEALSDELV 1240
QY 621 -----AFKNHKLKOLE-----VASAL--SHIAFNALDDNDGDQF 654
Db 1241 KSINNVDADAKKELKLEKMLQPGGYSNSQIEAMQSVKSALESYI--SASEATSTQEMN 1298
QY 655 NKVVVTHNSYALADGEHFIIVDPDKLSSTIVD-----LEKILKLEI-----DYSTLR 704
Db 1299 KQALVEAGTSLNWTDDQEKANEETKTSMYVVDVKYKEALEKVNABIDKYNKVQNDVPKYS 1358
QY 705 QTTQTQFRD--MTTAGKALLSKNLRQGEKQKFLQEAQFPLGRVLDLKAIAKALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQOKKKLMQ--EQAKLLKD--QIKSGNI-----TQYGIWTS 1404
QY 764 ATKNG--QLLERSINKAVLAYNNSAIKKNVNRLEKDLTLGLVEGK--GPLAQATMVOG 820
Db 1405 TSSGGTSPSTGSGYSKYSYNSAASKYNV-----DPALIAAVIQOESGFNAKARSGVG 1459
QY 821 VYLLKTPLPPEYIYGLNVYFDKSGKLIYALDMSDTIGEGOK--DAYGNPILNVDEBNEY 879
Db 1460 AMGLQMLPATAKSLGVNNAYDP-----YQNVMGTKYLAQOLEKEGG--NVEKALAY 1511
QY 880 HALAVATLADYEGLD-----IKTIL---NSKLSQTSIROVPTAAYHRAGIFOAION 928
Db 1512 NA--GPGNVIKYGGIPPEKFTQNVVKIMANYSKLSATS-----SIASY----- 1556
```

QY 929 AAABAEQLPKPQTHSEKSSSESANXDRGLQSNPKTNRGHRSAILPRTGSKG--SFVY 986
 Db 1557 -----TNSAFRVSSKYQGSGSLRSSP--HKGTDFAAKAGTAKLSQSGKV 1601
 QY 987 GILGYTSA 995
 Db 1602 QIAGYSKTA 1610

RESULT 10

PCT-US93-07261-11
 ; Sequence 11, Application PC/TUS9307261
 ; GENERAL INFORMATION:
 ; TITLE OF INVENTION: PEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John H. C. Blasdale
 ; STREET: One Giralda Farms
 ; CITY: Madison
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07940-1000
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh 6.0.5
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/07261
 ; FILING DATE: 19930805
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/927,531
 ; FILING DATE: 07-AUG-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Blasdale, John H. C.
 ; REGISTRATION NUMBER: 31,895
 ; REFERENCE/DOCKET NUMBER: DX0288K
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-822-7398
 ; TELEFAX: 201-822-7039
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1488 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US93-07261-11

Query Match 3.9%; Score 196.5; DB 5; Length 1588;
 Best Local Similarity 21.1%; Pred. No. 7.3e-05;
 Matches 209; Conservative 144; Mismatches 372; Indels 267; Gaps 53;
 QY 46 DYAGSS-GSKLKNINETSQVDD-TVTDLFSDKRTT-----PEKIKONLAKGPREQBL 96
 Db 611 EYDQSELAKGDIITKPHSVDEYDQTELAKGKEVTKNPHENLEYNEDTLAKG----- 664
 QY 97 KAVTENTESEKOITSGSQLESLSLNKTVPTSNWEICDFTTKGNTLVGLSKSGVEK 156
 Db 665 KEVTKNPHSVDEYDQSELAKGKD--ITNKPESVDEYDQTE-LAKGKEVTKNARENLEE 721
 QY 157 LSQTHLVLPQAADGTQLIVASFAFTPKKTATA---EYTSRAGEN-GEISQLDV-DG 211
 Db 722 YNETD-----LAGKKEVTKNARENLEYNEDTLAKGKEVTKNKAHENLEYNEDTLAKG 774
 QY 212 KEIINEG----BVFNSYLLKKVPTIGYKHIGQDAFVDNKNIAEYNLPESLETISDYAPA 267
 Db 775 KEVTKNKAHENLEYNEDTLAK-----GKE--VTNK--AHENLEYNED--DLAKG 818
 QY 268 HUALKQI--DLDPNLKAIGELAPF-DNQITGKLSLPROLMRLAERAFKSNHIKTTIFRGN 324
 Db 819 ----KEVTKNARENLEYNEDTLAKGKEVTKN-----ARENLEYNEDTLAKGK 863
 QY 325 SLKVTIGEASFQDNDLSQLMLPDGLEKIESEATGPNPGDDHYNRVVLMTKSGKNPSGLAT 384

Db 864 --EVTNKARENLEYNEDTLAKGKE--VTNKA---RENLEYNEDTL---AGKEVTNKA 914
 QY 385 ENTYNPDKSLWQSPSEIDYTKMLE-----EDFTYQKNSVTGFSNKGLOKQYKR 432
 Db 915 ENL-----EYNETDLAKGKEVTKNARENLEEYEEKDYMKNN--ELQNKGSGLKE 963
 QY 433 NKNLEIPKQHNGVVTITEIGDNAFNVDPONKTLRKYDLEEVKLTSTRKIGAPAFQSNNL 492
 Db 964 NAELKNKELRN-----KSGDGLKENAELKNKELNKGSDGLK-----ENAE 1005
 QY 493 KSFE-ASDDLEIEKEGAFMNNRIETLEKDKLVIGDAAPHINHIYAVLPESVQEIERS 551
 Db 1006 KKNELNKGSEGLKENAELKNK-----ELRNK-----GSEGLKEN-----AELKNK 1046
 QY 552 AFRQGANNLIFMGSKVTKLGBMAPLSNRLEHLDSEKQLTEIPVQAFSDNA-LK-EVL 609
 Db 1047 ELRNKGSSEG-----LKENAELKN-----KELQNKSGSEGLKENAELKNKEL 1086
 QY 610 LPASLKTIREBAFKKNHLKQLEVASALSHTAFNALDDNDGDEQFDNKVVVTKTHNSYALA 669
 Db 1087 QNKSGSEGLKENAEQKN--KELQ-----NKGSEGLKENAELKNKELRNKGS 1129
 QY 670 DGEHFIVDPDKLSSTIVDLBKILIRGLDYSTLROTTOTOFQDMTTAG-----KALLSK 724
 Db 1130 DGL-----KENAELKNKELRNKGSQ-----LKENAELKNKELRNKGSGLKENVVTN 1177
 QY 725 SNLRQGEKQKFLQBAQFPFLGRVDLDKAIKAKAKAL-----VTKKATKNQQLLRSI----- 775
 Db 1178 NDKNNDIQ-----NKLDSNKMKNKELNKLNDISNKMKNKELLNKLDSNEDMK 1226
 QY 776 NKAIVAYNNSAIKCANVKRLEKELDLTGL--VEGKGPLAQATWVGYYLLKTLPLPLPEY 833
 Db 1227 NKEILL--NKDIRNKDLKSGNMBEQONTGLKNTPSKQ-----QNTGLKNTP----- 1270
 QY 834 YIGLVYFDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDSDNBYHALAVATLADYBGL 893
 Db 1271 ----NERQONTG-----LAKTPEEQONTGLKNTPEEQONTG----- 1304
 QY 894 DIKTLNLSKLQSLTSIRQVPTAAYHRAIGIFQAIQNAABAEQLLPKPTHSEKSSSSSA 953
 Db 1305 -LKNTPNR-QONTGLKNTPEEQONTG-----LAKTPEEQONTGLKNTPEEQONTGLK 1358
 QY 954 NSKDRGLQS---NPKTNRGHRSAILPRTGSKG 982
 Db 1359 NAANKGOONTGLKNAANKGOONTGLKNTPSKG 1390

RESULT 11

PCT-US93-07261-16
 ; Sequence 16, Application PC/TUS9307261
 ; GENERAL INFORMATION:
 ; TITLE OF INVENTION: PEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John H. C. Blasdale
 ; STREET: One Giralda Farms
 ; CITY: Madison
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07940-1000
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh 6.0.5
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/07261
 ; FILING DATE: 19930805
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/927,531
 ; FILING DATE: 07-AUG-1992
 ; ATTORNEY/AGENT INFORMATION:

NAME: Blásdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0288K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-822-7398
TELEFAX: 201-822-7039
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1663 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
STRAIN: Malayan Camp
PCT-US93-07261-16

Query Match	3.9%;	Score 196.5;	DB 5;	Length 1663;
Best Local Similarity	21.1%;	Pred. No. 7.8e-05;		
Matches 209;	Conservative 144;	Mismatches 372;	Indels 267;	Gaps 53;
Qy	46	DYABSS--GSKLKLNETSGPVDD--TVDTLFSDKETT-----PEKIKONLAKGPRQEL	96	
Db	611	EYDQSELAKGDIYNKPHESSVDEYDQTELAKGKEVTNKPHEENLEEYNETDLAKG-----	664	
Qy	97	KAVTENTESEKQITSGSQLESLSINKTVPSTSNWEICDFITKGTNTVLGLSKSGVEK	156	
Db	665	KEVTNKPHESSVDEYDQSELAKGKD--ITNKPHESSVDEYDQTE--LAKGKEVTNKAARENLEE	721	
Qy	157	LSQTDHLVPSQAADGQTQLQVASFATPPDKKTAIA---EYTSRAGEN-GEISQLDV-DG	211	
Db	722	YNETD-----LAKGKEVTNKAARENLEEYNETDLAKGKEVTNKAHENLEEYNETDLAKG	774	
Qy	212	KEIINEG-----EVNSYLLKKVTIPTGYKHHGQDAFVDNKNIAEAVNLPESETISDYAPA	267	
Db	775	KEVTNKAHENLEEYNETDLAK-----GKZ--VTNK--AHENLEEYNET--DLAKG	818	
Qy	268	HLAKQOI--DLPDWLKAIAGELAFF-DNQITGKLSLPRQLMLRAERAFKSNHKTIEPRGN	324	
Db	819	----KEVTNKAARENLEEYNETDLAKGKEVTNK-----ARENLEEYNETDLAKGK	863	
Qy	325	SLKVIIGASQDNDLSQLMLPDGLEKIESEAFTPGCDHDDYNNRVLWTKSGNPSGLAT	384	
Db	864	--EVTNKAARENLEEYNETDLAKGKE--VTNKA---RENLEEYNETD---AKGKEVTNKAR	914	
Qy	385	ENTYVNPDKSLWQSPSIDYTKWLE-----EDFTYQKNSVTGFSNKGLOKVKR	432	
Db	915	ENL-----BEYNETDLAKGKEVTNKAARENLEEVEEKDYMKNN--ELQNKSGDGLKE	963	
Qy	433	NKMLEIPKHNGVTIITEIGNAFRVDFQNTLRKYDLEEVKVLPSSTRKTKGAFAPQSNNL	492	
Db	964	NABLKNKELRN-----KGSGLKENABLKNKELRNKSGSDGLK-----ENAE	1005	
Qy	493	KSPE-ASDLDLEEIKEGAFMNNRIETLELKKLVITIGDAAAPHINHIYAILVPESVQETGRS	551	
Db	1006	KNKELRNKSGEGLKENABLKNK-----ELRNK--GSEGLKEN-----AELKNK	1046	
Qy	552	AFRQNGANNLIFMGSKVKTILGEMAFPLSNRLEHLDLSBQKQITETPQVAFSDNA-LK-EVL	609	
Db	1047	ELRNKSGEG-----LKENAELKN-----KELQNKSGSEGLKENABLKNKEL	1086	
Qy	610	LPASLKTIRBEAFKKNHLKQLEVASALSHTAFNALDNDGDGQPDNKVVVKTTHNSVALA	669	
Db	1087	QNKSGEGLKENAEOKN--KELQ-----NKGSEGLKENABLKNKELRNKGS	1129	
Qy	670	DGEHFIVDPDKLSSTIVLEKILKLIBGLDYSTLRQTQTQTFRDMTAG-----KALLSK	724	
Db	1130	DGL-----KENAELKNKELRNKSGDG-----LKENAELKNKELRNKSGEGLKENVYTN	1177	
Qy	725	SNLRQGEKQKFLQEAQFQFLGRVLDKAIKAEKAL-----VTKKATQNGQLLERSI-	775	
Db	1178	NDLKNNDIQ-----NKLDSNKDMKNKELLNKKDI SNKOMKNKELNKLKOLSNBDMK	1226	

RESULT 12
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lyvin Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: staphylococcus epidermidis
US-09-134-001C-5080

Query Match	3.8%;	Score 195;	DB 3;	Length 3696;
Best Local Similarity	18.8%;	Pred. No. 0.00033;		
Matches	220;	Conservative 164;	Mismatches 478;	Indels 310; Gaps 45;
QY	10	LTLTVVVTHNQEVPSLAKPELIKQTQASSISIGADYAESSGSKLKNINETSGPVDVDTV	69	
Db	1082	ITHTNVVQKPPQARQALTAKTNEKQASINDNEGT--IBEKQAIQSLDANKNLADEQI	1139	
QY	70	TDLFSDKR-----TTPEKIKDNLAK-----GPREQBLKAVTENTSEKQ	108	
Db	1140	TOAASNQVNDNALNIGISNISKIOTWFTKKQARDQVQNFQBEAEELNSTPHATQDEKQ	1199	
QY	109	ITSGSOLEQKESLSLNTKVPSTSNWEICDFITPKG-----	143	
Db	1200	-DALTRLTOAKET-ALNDINQAOTNQVNDTALTSGIQNIQNTQVNRKKQEAKTINDIV	1257	
QY	144	-----NTLVLGSKSGVEKLSTQDHLVLPQO---AADGTQLIQV	178	
Db	1258	QQHKQIQONDDATTEEKEVANNLVNASQNV--ISKIDNATTNNQIDGIVSGRQGIN-	1314	
QY	179	ASFAFTP-----DRKTAIAEYTSRAGE-----NGEISQLDVGDKIIL-	215	
Db	1315	---AITPOTSIRKNAKNDIDIKAADKKIKIQRINDATDDEBIQEANRKIEAEAKDNQI	1371	
QY	216	-----HEGEVFNYSLLKKVITPYGKHIGODAFVDNK-----NIAEWNLPESL	258	
Db	1372	RNSTRQVNEAKTNGINKINIEITPATTVKSEARQA-VQNKANBQINHIQNTPATNEEKQ	1430	
QY	259	ETISDVAFAHALKQIDLPNLKAIGELAFPDNMOITGKLSLPQLM--RLAEARAFKSNHI	316	
Db	1431	EALNRYS-ELARVQRIQINAEHTTQGVKTIKDDAITSLSINQAUVKEKESARNAIEQAT	1489	
QY	317	KTIEFRGNSLUKIVGEASFQDNDDLSQMLPDGLBKIESEAFGTGNPGDDHYNRVLWTMKG	376	

```
Db 1490 QQTQFINNDNATDEEKEVANNLVATKQKSLDINSLS-----SNNDEVAKVAG 1540
Qy 377 KN-----PSGLATENTYVNPDKSLWQSPSPEDYTKWLEEDPTYQKNSVTGFSNKGLOKV 430
Db 1541 INEIANVLPATAVSKAKKIDIDKLAQTNQIQ-----THQ-TATTEKEAAIQLA 1590
Qy 431 KENKN---LEIPKQHNGVTIIEIGDNAPRNVDPQNTLRKYDLEBVKLPSTIRKIGA---484
Db 1591 NOKSNEARTAIQENSHNGVAQKNGIHEI-----ELVMPDAHKHSDAKOS 1637
Qy 485 ----PAFOSNNLKSFEASDDLEIEKEGAFMNNRIETLSLKKLVIGDAAFIHNIYAIV 540
Db 1638 IDNKYEQSNTINT--TPDATDEEKQAL-----DKLTKADAGY--NKVDQQAQ 1682
Qy 541 LPESVOETGRSFRONGANNLIFMGSKVKYTGEMAFPLNRHLDSQKQLTEIPVQAP 600
Db 1683 TNQVSDAKTEAI--DTITNI--QANVAKPSARVELDSKPEDL-----KRQINATP---1730
Qy 601 SNALKEVLLPASLTIREEAFKKNHLKQLEVASALSHTAFNALDDNGDQFQDNKVUVK 660
Db 1731 --NATEEB-----KODAIQRLNGKRDVKNLIN-----QQRDNEV---1764
Qy 661 THNSYALADGHEFVDPDKLSSTIVDL-----BKILKLIRGL 698
Db 1765 EQHKNIGLQLEETIHANPTKPSDALQELQTKFISQTELINNNKDATNEEKEAKELLEIS 1824
Qy 699 DYSTLRQTTQTQ-----PRDWTAGKALLSKSNLRQGEKQKPLQBAQPLGR- 745
Db 1825 KNTKTINQAQTNQVNDKNGMNEIATIPATTIKTDAKTAIDKKAEOQVTTINGNN 1884
Qy 746 --VDLDKALAK--AEKALVTKKATKNGQLLESINKA---VLANN---SAIKKANVKR 794
Db 1885 DATDEKAEARKLVEKAKIEAKSNTNSDTERVNGAKTNGLEKINNIPQSTQTKNAQ 1944
Qy 795 -----LEKELDLT-----GLVEGKGPLAQATMVGVYLLKTLPL 830
Db 1945 BINDKAQEQLIQNTTPDATEEKEQENATRVNAGLAQAIQINNASHSTQEVNESKTSIA 2004
Qy 831 PYYIYGLNVYFKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAVATLAD- 889
Db 2005 TIKSVQPNVI-----KPEATINSLTQEAANNQKTLIGDGNATDDEKEAAQLVTTQKLEQ 2059
Qy 890 ----VEGLDIKTILSKLSQTSIRQVPTAAVHRAGIFQAIONAA--AEAEQLLPKPGTH 943
Db 2060 IQKHESQDNQDQNVKQAQITAIKILINAHAKRQDAINILTLAESKKSDIRANQDATT 2119
Qy 944 SEKSSSESANSKDRGLQSNPKTNRGHSAIL 975
Db 2120 BEKTAIQSID--DTLAQARNNINGANTALV 2149
```

RESULT 13

```
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6
```

```
Query Match 3.8%; Score 192.5; DB 1; Length 2482;
Best Local Similarity 20.2%; Pred. No. 0.00027;
Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

Qy 6 KTVALTITVSVVTHNQEVFSLVKPEILKQTOASSISGADYAESGSKSLKINETSPV 65
Db 1341 KTTALDQSEKMKETQEL-----ESHQSECLHCIOVAEAEVKEKTELLQT---L 1387
Qy 66 DDTVDLPDSDKRTTPEKI---KDNLAGPRPEQLKAVTENTESEKQITSGSQLEQSKES 121
Db 1388 SSDVSELLKDKTHLOEKLOSLEKSOALSUTKCELENQIAQLNKEKELL-----VKES 1440
Qy 122 LSLNKTVPSTSNWICD-----FITKGNLTVLGS-----KSGVEKLS---QTD 161
Db 1441 ESLQARL--SESDYEKLVNKALEALVEKGEFALRLSSQTEVHQLRGIEKLRVRIEAD 1499
Qy 162 -----HLV--LPSQAADGTQLIQVASFAPTDKKTATAEYTSRAGENCEISQLDVG---211
Db 1500 EKKQLHIAEKLKREKRENDLSL-----KDKVENLERELQMESENEQELVILDAENSKA 1550
Qy 212 -----KEIINEGEVFNYSLLKKVTIPTGYKHI-----GQ-----DAF 243
Db 1551 EVETLKQTQIEBMAASLKVFE---IDLVTLRSEKENLTKIOEQKQQLSELKLLSSPKSL 1607
Qy 244 VDNKNIAEVNLPESETISDYAFALHAKQIDLPDNLKAIGELAFPDNQITGKLSLPRQL 303
Db 1608 LEEKEQAEIQIKERSKTAVEM-----LQNQLKEL-----NEAVALCGDQEI 1649
Qy 304 MRLAERAPKSNHIKTIEFRGNSLKVIGEASFQDNDLSQMLPDLGLEKIESEAFNGPDD 363
Db 1650 MKATEQSLDPPIEBHQLR--NSIEKL--RARLEADEKKQLCVLQQLKSEHHA-----D 1700
Qy 364 HNNNRV-----VLMTKSGKNPGLATENTYVNPDKSLWQSPSPEDYTKWLEEDPTYQKN 417
Db 1701 LLKGVENLERELSEIARTNQEHAAEAENS-----KGEVETLKAKIEGTQ---1746
Qy 418 SVTGFSNKGLO-----KVKRNKNLEIPKQHNGVTIIEIGDNAPRNVDFQNTLRKYDYL 470
Db 1747 ----SLRGLDLVVVTIRSEKENLTNELQKEQERISELEIINSSPENI--LQEKQEKVQM 1800
Qy 471 EEVKLPSTIRKIGAFQSNLKSFEASDDLEEK--EGAFMNNRIETLEKDKLVITGDA 529
Db 1801 KEKS--STAMEMLQTLQKELNVERVAALHNDQEAQKAEQNLSSQVECELEKQAQLQGLD 1858
Qy 530 AFHINHIAIVLPESV-----QEIGRSAPFRONGANNLIFMGSKVKTLGEMAFPLNRHL 585
Db 1859 EAKNNY----IVLQSVNGLIQEV-----EDGKQKL-----EKKDEISRLKNQIQ---1900
Qy 586 LSEQKQLTEIPVQAFSDNAL--KEVLLPASLTKTIRREAFKKNHLKQLEVASALSHTAFNAL 644
Db 1901 --DQEQVLVSKLSQVEGEHQLWKEQNL--ELRNLTVLEFQKIQVLQSKNAS-----L 1947
```


QY 645 DNDGDEQFDNKKVVKTHNSYALADGHEFIVDPKLSSTIVDLKILKIEGLDYSLR 704
Db 1948 QD-----TLEVLQSSYKNIENLELTMDKMSF-----VEKVNKM-----TAK 1985
QY 705 QT-TOTOFDMTMTAGKALLSKNSLRQGEKQKFLQEAQFPLGRVLDLKAIAKAELVTK 763
Db 1986 ETELQREHMHM--AQKTAELQBEL--SGEKNRLAGELQLLLEI-----K 2026
QY 764 ATKNGQLLERSINKAVLAYNNSAIKKA-----NVKRLKEKELDLITGLVGEKGPQAQATWVQ 819
Db 2027 SSKD-QLKE-----LTLNSELKSLDCKMHQDQVEKGVREBIAEVLRLHEAEK 2078
QY 820 GYVLLKTPPLPYEYIGLVNVPDK-----SGKLIYALDMSDTIG 858
Db 2079 QALLLDTN---KQYEVEIQTREKLTKEECLSSQKLEIDLKSSKEELNNSLKATQTIL 2135
QY 859 EQQKDA-----YGNPILNVDDNNGYHALAVATLADYEGLDIKITLNSKLSOL----- 906
Db 2136 BELKTKMDNLKYVNLKKNERAGQKMKLLKSCQLE--BEKEILOKELSQLQAQEK 2193
QY 907 -----TSIRQVPTAAVHRAGIFQAIONAAAEAEQLPKPG-----THSEKSSSESAN 954
Db 2194 QKTGTVMTKVDELTEIKE---LKETLEKTKDEADYLDKYCSLLISHEKLEKAKEMLE 2250
QY 955 SKDRGLQSNPKTNRGRHSAIL 975
Db 2251 TQVAHLCSQSQSDSRGSPLL 2271

RESULT 14

US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN

US-08-353-700-1

Query Match 3.8%; Score 191.5; DB 1; Length 3248;
Best Local Similarity 20.1%; Pred. No. 0.00047;
Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;
QY 6 KTVALTLTTSVVTHNQVFLSVKPEILKQTOQASSISGADYAESGSKLKINFTSGPV 65
Db 2069 KTTALDQLSEKMKETQEL-----ESHQSECLHCIOVAEAVKEKTELLQT---L 2115
QY 66 DDTYTDLFSDKRTTPEKI-----KONLAKGPBEQELKAVTENTSEKQITSGSQLSQSKES 121
Db 2116 SSDVSELKDKTHLQELQSLQSEKQALSHTKCELENOIAQNKSEKELL-----VKES 2168
QY 122 LSLNKTVPSTGNWEICD-----FITKGNLTVLGLS-----KSGVEKLS---QTD 161
Db 2169 ESLQARL--SESDYEKLVNKALEAALVEKGFALRLSSTQBEVHQLRGIEKLRVRIEAD 2227
QY 162 -----HLV--LPSQAADGTQLIQVASFAPTPDKTAIAEYTSRAGENGIEISOLDVDG--- 211
Db 2228 EKKQLHIAEKLKERENDSL-----KDKVENLERELQMSEENQELVILDAENSKA 2278
QY 212 -----KEIINEGEVFNYSLLKKVTIPTGYKHI-----CQ-----DAF 243
Db 2279 EVETLKTQIEEMARSLKIFE---LDLVTLRSEKENLTQIOEKQQLSELDKLSSFKSL 2335
QY 244 VDNKNIAEVLNLPESLETISDYAFALHALKQIDLPNLKAIGELAFDFNQITGKLSLPRQL 303
Db 2336 LEEKEQAEIQKERSKTAVEM-----LQNLKEL-----NEVAALCGDOEI 2377
QY 304 MRLAERAFKSNHIKTIEPRGNSLVIGEASFDQNDLSQLMLPDGLEKIESEAFTPNGDD 363
Db 2378 MKATEQSLDPPIESHEQLR-NSIEKL--PARLEADEKKQLCVLQQLKESHEHA-----D 2428
QY 364 HNNRV-----VLWTKSGKNPSGLATENTYVNDPKSLWQSPEDITYKWLBEEDTYQKN 417
Db 2429 LLKGRVENLERELEIARTNQEHAALAEANS-----KGEVETLKAKIEGTMQ--- 2474
QY 418 SVTGFSNKGLQ-----KVKRNKNLEIPKQHNGVITTEIGDNAPRVDFQNKTLRKVDL 470
Db 2475 -----SURGLELDVVVTRSEKENLTNELOKQERISELEIINSSPENT--LOEKEQEKVQM 2528
QY 471 BEVKLPSTIRKIGAFQFOSNNLKSFEAGDDLEETK--EGAFMNNRLETLELCKLVITGDA 529
Db 2529 KEKS--STAMEMLQTLQKELNERNVAALHNDQEAACKAEQNLSSQVECELEKQAQLQGLD 2586
QY 530 AFHINHIYAIVLPESV-----QEIGRSAPFRONGANNLIFMGSKVKVTGLGEMAFLSNRLEHL 585
Db 2587 EAKNYY---IVLQSSVKGLIQEV-----EDGKQKL-----EKKDEISRLKNQIQ--- 2628
QY 586 LSEQKOLTEIPVOAFSDNAL--KEVLLPASLKTIRBEAFKQKHLKQLEVASALSHTAFNAL 644
Db 2629 --DQEQLVSKLSQVEGEHQLWKEQNL--ELRNLATVELEQKIQVLQSKNAS-----L 2675
QY 645 DNDGDEQFDNKKVVKTHNSYALADGHEFIVDPKLSSTIVDLKILKIEGLDYSLR 704
Db 2676 QD-----TLEVLQSSYKNIENLELTMDKMSF-----VEKVNKM-----TAK 2713
QY 705 QT-TOTOFDMTMTAGKALLSKNSLRQGEKQKFLQEAQFPLGRVLDLKAIAKAELVTK 763
Db 2714 ETELQREHMHM--AQKTAELQBEL--SGEKNRLAGELQLLLEI-----K 2754
QY 764 ATKNGQLLERSINKAVLAYNNSAIKKA-----NVKRLKEKELDLITGLVGEKGPQAQATWVQ 819
Db 2755 SSKD-QLKE-----LTLNSELKSLDCKMHQDQVEKGVREBIAEVLRLHEAEK 2806
QY 820 GYVLLKTPPLPYEYIGLVNVPDK-----SGKLIYALDMSDTIG 858
Db 2807 QALLLDTN---KQYEVEIQTREKLTKEECLSSQKLEIDLKSSKEELNNSLKATQTIL 2863
QY 859 EQQKDA-----YGNPILNVDDNNGYHALAVATLADYEGLDIKITLNSKLSOL----- 906
Db 2864 BELKTKMDNLKYVNLKKNERAGQKMKLLKSCQLE--BEKEILOKELSQLQAQEK 2921

QY 907 -----TSIRQVPTAAYHRAGIFQAIQNAARAEQLLPKG-----THSEKSSSESAN 954
Db 2922 QKTQVMDTKVDELTEIKE---LKETLEETKTEADYLDKYCSLLISHKELEKAKEMLE 2978
QY 955 SKDRGLQSNPKTNRGRHSAIL 975
Db 2979 TQVAHLCSQSQSKQDSRGSPLL 2999

RESULT 15
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3448 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 3.8%; Score 191.5; DB 5; Length 3248;
Best Local Similarity 20.1%; Pred. No. 0.00047;
Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;

QY 6 KTVALLTLTVSVVTHNQVFLVKEPILKQTAQASSISGADYAESGSKLKNINETSQPV 65
Db 2069 KTTALDQSEKXKKEKTSGL-----ESHQSECLHCIQVAEAVREKTELLQTL 2115

QY 66 DDTVTDLFSKRTTPEKI-----KONLAKGPREQELKAVTENTSEKQITSGSQLEQSKES 121
Db 2116 SSDVSELKDKTHLQELQSLKESQALSILTKCELENQIAQLNKEKELL-----VKES 2168

QY 122 LSLAKTVSTNWEICD-----FITKGNVLVGLS-----KSGVEKLS---QTD 161
Db 2169 ESLOQARL-SESDYKLNYSKALEALVKEGFEALFLSQTBEVHQLRGIEKLRVRIEAD 2227
QY 162 -----HLV--LPSQAADGTQLIQVASFPTPKKTAIABYTSRAGENGIEISQLVDG--- 211

Db 2228 EKQHLIAEKLKERENDSL-----KDKVENLERELQMSSENOELVILDAENSKA 2278
QY 212 -----KEINIEGVPNSYLLKKVTIPTGYKHI-----GO-----DAF 243
Db 2279 EVETLKTQIEEMARSLKIFE---LDLVTRSEKENLTKIQEQKQQLSELDKLLSSFKSL 2335
QY 244 VDNKNIAEAVNLPESLETISDYAPFAHLKQIDLPDNLKAIGELAFPNQITGKLSLPROL 303
Db 2336 LEEKEQAEIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 2377
QY 304 MRLAERAFKSNHIKTIIEFRGNLSKVIGEASQDNDLSQLMLPDGLEKIESAFTGNPGDD 363
Db 2378 MKATEQSLDPPIEBHQLR-NSIEKL-RARLEADEKKQLCVLQQLKSEHHA-----D 2428
QY 364 HYNRV-----VLWTKSGKNPSGLATENTYVNPDKSLWQSPRIDTYKMLEEFTYQKN 417
Db 2429 LLKGRVENLERELIARTNOEHALEANS-----KGEVETLKAKIEGTMQ--- 2474
QY 418 SVTGFSGKGLQ-----KVRKNKRLIPKOHNGVTITEIGDNAPRVNDFONKTLRYKDL 470
Db 2475 -----SURGLELDVVTRSEKENLTNLSQKEQERISELEIINSSFENI-LOEKEQEKVQM 2528
QY 471 BEVKLPSTIRKIGAFQSNLKSFEASDDLEBIK-EGAFMNNRIETLEKDKLVTTIGDA 529
Db 2529 KEKS--STAMEMLQTLKELNERNVAALHNDQEAACKAEQNLSSQVECLEKALQGLD 2586
QY 530 AFHINHIYAVLPESV-----QETGRSAFRONGANNLIFMGSKVKTILGEMAPLSNRLEHD 585
Db 2587 EAKNY---IVLQSSVKGLIQEV-----EDGKQKL-----EKDDEISRLKQIQ--- 2628
QY 586 LSEOKQLTEIPVOAFSDNAL-KEVLLPASLKTIRREAFKKNHKLQLEVASALSHPAL 644
Db 2629 --DOEQLVSKLSQVEGSHQWQNL--ELRNLTVELEQKIQLVQSKNAS-----L 2675
QY 645 DDNDGEQFONKVVVTKHNSYALADGEHFTVDPDKLSSTTVTDLEKILKLTIEGLDYSTLR 704
Db 2676 QD-----TLEVLQSSYKNLENELELTQMDKMSF---VEKNKM-----TAK 2713
QY 705 QT-TQTFQDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTYK 763
Db 2714 ETELQREMHEN--AQKTAELQEEL-SGEKNRLAGELQLLLEFI-----K 2754
QY 764 ATKNGQLLERSINKAVLAYNNSAIKKA-----NVKRLKELDLTLGLVEGKGPLAQATMVQ 819
Db 2755 SSKD-QLKE-----LTLENSELUKSLDCHMKHQQVEKGVREIRIAYQRLHEAEKHH 2806
QY 820 GVVLLKTPPLPEYIIGLVNVPFK-----SGKLIYALDMSDTIG 858
Db 2807 QALLLDTN---KQYEVEIQTYREKLTSEBECLSSQKLEIDLKSSKEELNLSKATTOIL 2863
QY 859 EGQKDA-----YGNPILNVDENEGYHALAVATLADYEGLDIDITILNSKLSQL----- 906
Db 2864 EELKTKMDNLKVVNQLKKNERAGQKMKLLIKCKQLE--EKEIILQKELSOLQAAQEK 2921
QY 907 -----TSIRQVPTAAYHRAGIFQAIQNAARAEQLLPKG-----THSEKSSSESAN 954
Db 2922 QKTQVMDTKVDELTEIKE---LKETLEETKTEADYLDKYCSLLISHKELEKAKEMLE 2978
QY 955 SKDRGLQSNPKTNRGRHSAIL 975
Db 2979 TQVAHLCSQSQSKQDSRGSPLL 2999

RESULT 16
US-09-538-092-1154
; Sequence 1154, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092

```

; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqformatter Version 0.9
; SEQ ID NO 1154
; LENGTH: 3210
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P49454
US-09-538-092-1154

Query Match          3.7%; Score 187.5; DB 4; Length 3210;
Best Local Similarity 20.1%; Pred. No. 0.00089;
Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;

QY      6 KTVALTITVSVVTHNQEVFSLVKEPILKQTQASSISGADYAESSGKSKLKINETS GPV 65
DB      2069 KTTALDQLSEKWKETQEL-----ESHQSECLHICIQVAEAEVKEKTELLQT---L 2115

QY      66 DDTVDFDSDKRTTPEKI----KDNIAKGPREGQELKAVTENTESEKQITSGSQLEQSKES 121
DB      2116 SSDVSELLKDKTHLQEKLSQKDSQALSITKCELENQALNKEKELL-----VKES 2168

QY      122 LSLINKTVPSTSNWEICD-----PITKGNLTVLGLS-----KSGVEKLS---QPD 161
DB      2169 ESIQARL--SESDYEKLNVSKALEAALVEKGEFALRLSSQTQEEVHQLRRGIEKLRVRIEAD 2227

QY      162 -----HLV--LPSQADGTQLQVASFAPTPDKKTAIAEYTSRAGENGEISQLDVDG--- 211
DB      2228 EKKQLHIAEKLKERENDSL-----KDKVENLERELQWSEENQELVILDAENSKA 2278

QY      212 -----KEIINEGEVNSVLLKKVTIPTGYKHI-----GO-----DAF 243
DB      2279 EVETLTKTOIEEMARSLKVEF---LDLVTLRSEKENLTQIOEKQQLSGLDKLISFSKSL 2335

QY      244 VNKNIAEVLNLPESLETISDYAFAPHALKQIDLPDLNKAIGELAFPDNQITGKLSPLROL 303
DB      2336 LSEKEQARTQIKESKTAIVEM-----LQNQLKEL-----NEAVALCGDQEI 2377

QY      304 MRLAERAPKSNHIKTIIEFGNSLKVIGEASFQNDLSQLMPLDGLKIESEAFNGPQGD 363
DB      2378 MKATEQSLDPPIEEBEHLR--NSIEKL--RARLEADEKKQLCVLQOLKESHHH-----D 2428

QY      364 HYNRRV-----VLWTKSGKNPGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417
DB      2429 LLKGRVENLERELEIARTQEHAALEAENS-----KGEVETLKAKLEGMTQ--- 2474

QY      418 SVTGFSNKGLO-----KVKRNKNLEIPKQNGVVTITIGDNAPRVDFQNKTLRKYDL 470
DB      2475 -----SLRGLEDVVTIRSEKEDLTWELQEKERISELEI INSSPENI--LQEKQEKVQM 2528

QY      471 EVKLPSTTRKIGAFQFQNNLKSFEASDDELEIK--EGAFPMNRITETLEKDKLVITGDA 529
DB      2529 KEKS--STAMEMLTQTLKELNERRAALHNDQEAACKAQNLSSQVECLRELAQLLQGLD 2586

QY      530 AFPHINHIYAILVPESV---QETGRSAPFGNGANNLI FMGSKVKTILGENAFISNRLHLD 585
DB      2587 EAKNNY---IVLOSSVNGLIQEV-----EDGQKL-----EKDDEEISRLKNQIQ--- 2628

QY      586 LSEQKQLTEIPVQAFSDNAL--REVLIPASLKTIREEAFKKNHLKQLEVASALSIAFNAL 644
DB      2629 --DQEQVLVSKLSQVEGEHLWKEQNL--ELRNLITVELEQKIQLVQSKNAS-----L 2675

QY      645 DNDGDGEQFDKVVVTHNSYALADGEHPIVDPPDKLSTIVDLEKILKIEGLDYSTLR 704
DB      2676 QD-----TLEVLQSSYKNLENELELTAKMDKMSF-----VEKVNQK-----TAK 2713

QY      705 QT-TQTQFRDMTTAGKALLSKSNLRQGEKQKQFLOBAQPFPLGRVLDLKAIAKAELVTKK 763
DB      2714 ETELQREMHM--AQKTAELQEEL--SGEKNRLAGELQLLLLEI-----K 2754

QY      764 ATYNGQLLERSINKAVLAYNNSAIKKA-----NVKLEKELDLTLTGLVGEKGPPLAQATWVQ 819
DB      2755 SSKD-QLKE-----LTLENSELKSLDCMKHQVKEKGVREIEIARYQLRLHEAEKKH 2806

QY      820 GVVLLKTLPLPEYVIGLVNVPDK-----SGKLIYALDMSDDTIG 858
DB      2807 QALLLDTN---KQYEVETQTYREKLTSKEECLSSQKLEIDLKSKSEELNLSKATQIL 2863

QY      859 EQQKDA-----YGNPILNVDEDNQYHALAVATLADYEGLDITILNSKLSQL----- 906
DB      2864 BELKTKMDNLKYNVQLKKENERAQGMKLLIKSKCKQLE--BEKELQKELSQLQAAQEK 2921

QY      907 -----TSIRQVPTAAYHRAGIFQATQNAAAAEQLLPKG-----THSEKSSSESAN 954
DB      2922 QKTGTVMDTKVDELATTEIKE---LKETLEEKTKEADEVLDKYCSLLISHEKLEKAKEMLE 2978

QY      955 SKDRGLQSNPKTNRGRHSAIL 975
DB      2979 TOVAHLCSQQSKQDSRGRSPLL 2999

RESULT 17
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match          3.6%; Score 185; DB 4; Length 3878;
Best Local Similarity 19.5%; Pred. No. 0.0018;
Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

QY      7 TVALTLTTVSVVTHNQEVFSLVKEPILKQT--QASSISGADYAESSGKSKLKINETS GP 64
DB      94 TIMRTLHSGEITSHQEG--FSVELESEISTTADDCSEVNGCSFVVRTGKPTNLLREERG 152

QY      65 VDDTVTDL--FSDKRTTPEKIDKNLAKPREQELKAVTENTESEKQITSGSQ--LEOSKESL 122
DB      153 VDDSYSEQAQSDSPTHLEMBSELA--GKQHEIBEINRELE--EMRVTYTGEGLQQLQEFPE 209

QY      123 SLNK-----TVPSTNWEICDPITKGNLTVLGLSKSGVEKLSQTDHLVLPDQADGTQLIQ 177
DB      210 AAIKQDGGIITOLTANLQOAR--REKDETM---REFLELTSQSKLQIQFOQLQOASETLR 264

QY      178 VASFAFTP---DKTATAEYTSRAGENGEISQLDQVDGKEIINEGEVEN---SYLLKKVYT 230
DB      265 NSTHSTAADLLQAKQIILTHQOQLEEQDHLE-DYQKKK-----EDFTMQISFLQEKIK 318

QY      231 IPTGYKHICQDAFVNKNIAEVLNLPESLETISDYAFAPHALKQIDLPDLN-----KAIGEL 286
DB      319 V-----YEMEQDKVNSKEEIOEKETIEBELNTKIIIEEKKTELKOKLTADKLLGEL 374

QY      287 AFFDNNQITG-----KLSLP-----RQMLRAERAFKSNHIKTIIEFGNS 325
DB      375 ---QBQIVQKNQEIKNMKLELTNSKQKERSSEIKQLMGTVLEQLKRNH--KDSQF----- 426
```



```
Db 208 NTGALTDTSQYQFKQ-----GLTKLNDYTPHNOIVNFEN--TSLETIDNYVTADSWYRP 260
Qy 269 -----LALKQIDL-----PDNLKAIGELAFPDNQITGKLSLPRQLMLAEBA 310
Db 261 KOILKNGKWTWASSEDRLPLMSWMPDKQTOIAYLNYMNQOGLG-----TGENY 310
Qy 311 FKSNIHKITIEFRGNSLKVGEASFQNDLSQLM-----LPDGLKIESEAFPTGNP 360
Db 311 TADSSQESLNLAQTQVQKIEKISQTOQWLRLDIINSFVKTPQPNWNSQTESDTSAGE- 369
Qy 361 GDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESP--EIDYTKWLEBDFTYQKNS 418
Db 370 -KDHLOGGALLYSNSDK-----KNAINHLSIL-----EAWSDNDPOY---NKDTKGAQL 413
Qy 419 VTGFSNKGLOKVKRNKNLEIPKQHNGVTITEIGDNAFRNVDPNKTLRKDYDEEVKLPT 478
Db 414 -----SSGGYDFLLAN-----DIDNSNPVVOAQLN----- 439
Qy 479 IRKIGAFQSNLKSFEASD-----DLEEKGAFFMNNRIETLEKDKLVLTIGDAAPHINH 535
Db 440 -----WLHYLMNYGSIIVANDPEANFDGVRVDVNDVNNADLLQI-----ASDYLKA 484
Qy 536 IYAIVLPSVQBSIGRFAFRONGANNLIIPGSKVKTLGEMAFLSNRLEHLDLSEQKLTBI 595
Db 485 HYGVDKSE-----KNAINHLSIL-----EAWSDNDPOY---NKDTKGAQL 521
Qy 596 PQVAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSASHIAPNALDDNDGD---E 651
Db 522 PI-----DNKRLSLLYALTRPLEKDSNKN-----EIRSGLEPVITNSLNRSABGKNS 572
Qy 652 QPDNKVVVKTTHNSY--ALADGEHFIWDP--DKLSSTIVDLKILKIEGLDYSTLRQTT 707
Db 573 RMANIYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFILDELKQAFKIYNEDMRQAKKKT 632
Qy 708 QTOFRDMMTAGKALLSK-----SNLRQGEKQKFLQEAQFLGRVLDLKAIAKALVTCK 763
Db 633 QS-----NIPTAYALMLSNKDSITRLYYGD--MYSDDQYMATKSPYYDAIDTLKARI--K 685
Qy 764 ATKNGOLLERSI---NKAVLAYNNSAIKKNVKNLEKELDLTLGLVEGKPLAQATWVG 820
Db 686 YAAGQDMKITVYEGDKSHMDYTGVLTS-----VRYGTGANEATDQGEATKTQG 737
Qy 821 VYLLKTPPLPEYIIGLVVFKSGKLIYALDMSDTIGEGQKDAYGNPILNVDENEGYH 880
Db 738 MAVITSNP-----SLKINGQDVIV-----NWGAHKNQOYRPLILLTTKGLTSY 783
Qy 881 ALAVATLADY-----EGLDIKTILNSKLSQTSIRQVPTAAYHRAGIFQAIQNA 929
Db 784 TSAAAKSLYRKTNDRKGLVFDASDIQGYLNPQVSGYLAV-WVPVCGASNDQVVRVAASNK 842
Qy 930 ABAEQLLPKPGTHSEKSSSES-----ANSKDRGLQSNPKTNR 968
Db 843 ANATQVY-----ESSALDSQLIYEGFSNFQDFVTKDSYTNK 881
```

RESULT 25

```
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
```

```
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6
```

Query Match 3.5%; Score 176; DB 4; Length 1430;

```
Best Local Similarity 17.9%; Pred. No. 0.0017;
Matches 191; Conservative 180; Mismatches 401; Indels 294; Gaps 43;
```

```
Qy 1 MKKHLKTLVALTLTTVSVVTHNQEVFSLVKPELIPKOTQASSISGADYAESGSKLKINE 60
Db 12 VKKHVVTVAVASGLITLGT-----TTLGSSVSAETEQTQS----- 46
Qy 61 TSGPVDVTVDLFSDKRTTPEKIKNLAKGPREQBELKAVTENTESEKQI--TSGSOLEQSK 119
Db 47 -----DKVVTQKSEDDKAASESSQTDAPKTKQAQ-----TEQTAQSOANVADTSTSIK 96
Qy 120 ELSLANKTVPSTSNWEICDFITKGNLTVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVA 179
Db 97 ETPSQNIITTOANS-----DDKTVTNTKSEEAQTSEERTKQSE-----AQTTASSQAALTOA 147
Qy 180 SFAFPTDKKTAIAB-----YTSRAGENGESISOLDVDVKIIEI--- 215
Db 148 KAEITKQQTAAQENKNPVDLAAIPNVQIDGKYIYIGSDGQPKNFALTNNKVLVYFDPK 207
Qy 216 NEGEVENS--YLLKKVTTIPTGYKHIGQDAFVDNKNIAEVLNPESLETISDYAFAH----- 268
Db 208 NTGALTDTSQYQFKQ-----GLTKLNDYTPHNOIVNFEN--TSLETIDNYVTADSWYRP 260
Qy 269 -----LALKQIDL-----PDNLKAIGELAFPDNQITGKLSLPRQLMLAEBA 310
Db 261 KOILKNGKWTWASSEDRLPLMSWMPDKQTOIAYLNYMNQOGLG-----TGENY 310
Qy 311 FKSNIHKITIEFRGNSLKVGEASFQNDLSQLM-----LPDGLKIESEAFPTGNP 360
Db 311 TADSSQESLNLAQTQVQKIEKISQTOQWLRLDIINSFVKTPQPNWNSQTESDTSAGE- 369
Qy 361 GDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESP--EIDYTKWLEBDFTYQKNS 418
Db 370 -KDHLOGGALLYSNSDK-----KNAINHLSIL-----EAWSDNDPOY---NKDTKGAQL 413
Qy 419 VTGFSNKGLOKVKRNKNLEIPKQHNGVTITEIGDNAFRNVDPNKTLRKDYDEEVKLPT 478
Db 414 -----SSGGYDFLLAN-----DIDNSNPVVOAQLN----- 439
Qy 479 IRKIGAFQSNLKSFEASD-----DLEEKGAFFMNNRIETLEKDKLVLTIGDAAPHINH 535
Db 440 -----WLHYLMNYGSIIVANDPEANFDGVRVDVNDVNNADLLQI-----ASDYLKA 484
Qy 536 IYAIVLPSVQBSIGRFAFRONGANNLIIPGSKVKTLGEMAFLSNRLEHLDLSEQKLTBI 595
Db 485 HYGVDKSE-----KNAINHLSIL-----EAWSDNDPOY---NKDTKGAQL 521
Qy 596 PQVAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSASHIAPNALDDNDGD---E 651
Db 522 PI-----DNKRLSLLYALTRPLEKDSNKN-----EIRSGLEPVITNSLNRSABGKNS 572
Qy 652 QPDNKVVVKTTHNSY--ALADGEHFIWDP--DKLSSTIVDLKILKIEGLDYSTLRQTT 707
Db 573 RMANIYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFILDELKQAFKIYNEDMRQAKKKT 632
Qy 708 QTOFRDMMTAGKALLSK-----SNLRQGEKQKFLQEAQFLGRVLDLKAIAKALVTCK 763
```

Db 633 QS---NIPYALMLSNKDSITRLYYGD--MYSDDGQYMATKSPYYDAIDTLTKARI--K 685
Qy 764 ATKNGOLLERSI---NKAVLAYNNNAIKKANYRLEKELDLTLTGVLVEGKGPLAQATWVG 820
Db 686 YAAGQDMKITTYVEGDKSHMDWDTYGLVTS-----VRYGTGANEATDQGEATKTQG 737
Qy 821 VYLLATPLPLPYIYGLNVPYDKSKLIYALDMSDTIGEGQKDAYGNPLNVDNEGVH 880
Db 738 MAVITSNNP-----SUKLQNDKVIV-----NMGAHHKQNYRPLLLTTKDGGLTSY 783
Qy 881 ALAVATLADY-----EGLDIKITLSKLSQTSIROVPTAAVHRAGIFAQIQA 929
Db 784 TSDAAKSLYRKTNDKGLVFPDASDIQGYLNPQVSGYLAV-WVPYGASDNQDVRVAASNK 842
Qy 930 AAEASQLLPKPTHSEKSSSSS-----ANSKDRGLQSNPKTNR 968
Db 843 ANATQVY-----ESSSALDSQLIYBGFNSFQDPVTKDSYTNK 881

RESULT 26
US-09-248-796A-20284
; Sequence 20284, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20284
; LENGTH: 1078
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20284

Query Match 3.5%; Score 175.5; DB 4; Length 1078;
Best Local Similarity 20.9%; Pred. No. 0.0012;
Matches 185; Conservative 142; Mismatches 326; Indels 231; Gaps 42;
Qy 9 ALTLTVSVVTHNRVVF-----SLVGPILKQTOASSISGADYAESGSKSLKIN-ET 61
Db 308 ALTLTFSQDATGLPTVYLTVDVYVNLVKDNFLR-----IKRALPHDPSSBPRGKISYEL 360
Qy 62 SGPVDDTVTDLPSDKRTTPEKIKDN-----LAKGPREQELKAVTEN 102
Db 361 PEELDTKVNLTKELQTEKENAESNDKELNEKIEKLTNLSTKLTKEKDEKEQELAKIQED 420
Qy 103 TS--SEKQITSSQLEQSKESLSLNKTVFSTSNWEICDPITKGTNLVGLSKSGVKLSQT 160
Db 421 HKLLNEKFLVTNLSAMKASKKEFETASQYKQKLEQALKKGNTSESTLKQLEKLDST 480
Qy 161 D-----HL-VLPQSAADGTQLIQ-----VASFAFTPDKKTAAIABYT 195
Db 481 EQAKKLEDDGINMTDRDLFLHKLKASEAE--TQIKQREREPKNTLYELENKKDYELQIN 538
Qy 196 SRAGENGEISQ-----LDVDGK-----EINEGEVFNYSLLKKVTIPTGYKH 237
Db 539 NLNKSNNPEFKKINELSKIESLTEDNKSNAKQLEBKLRDTEENHEHMLDKLRAS----- 594
Qy 238 IQGDAFVDMKNIAE---VNLPSLETISDYAPAHALAQIDLDPNL-KAIGLAFPDNOI 293
Db 595 VAYNDLKKAKSSEETVAKSELETLSKI-----DNLEKELKEQQSKNEL 642
Qy 294 TKLSLPLQMLAERAFK--SNHIKTIEFRGNSLKVTGEASFQDNLSQLMLPDGLEKI 351
Db 643 EQOL---QNITDSTNEKFKLEDELKSIK-----KSNKEISSQNSLIQ-----KLEKT 688

Qy 352 ESEFTGNPGDDHYNRNVLTWTKSGKNPGLATENTYVNPDKSLWQSPSEIDYTKWLEED 411
Db 689 EKOL-----QAKDEIDKLAETK-----SNIDNUNSEISSLSQSKLKEA-----BES 730
Qy 412 FTYQKNSVTGFSNKGLOKVK-----RNKNLEIPKQHGNGVTIPEIGDNAPRNVDFQNKTLRK 467
Db 731 HSSTKDEHSSLS-ENLKKLKEEYENTKTMIAK-----LSAKIBEHKKATDEIETKTKHI 784
Qy 468 YDLBEVKLPSTIRKIGAFAPOSNNLKS-----PEASDDLEBI-----KEGAFMNNRIETL 517
Db 785 TDLOE-----EHAQKQSPESERNDIKSNLDEANKELSDNREKLSNLEKEKTELANKLKTQ 840
Qy 518 ELKDKLVTIGDAAHINHIYAIVLPESVQETIGRAFRQNGANNLIFMGSKVKTIGEMAF 577
Db 841 BEK---ISDLTSETVAISEDKSKLKHIEDLKR-----EKIKL---ETTLK 880
Qy 578 SNRLEHLDLSRQKQLTEIPVQAFSDNALKEVLLPASLKTIRE-----EAPKK 624
Db 881 ENBESTMFEKKEQLQVV-----NDKCKE--LEACLKLTETKEKEINDLIRKLEAKS 930
Qy 625 NHLKQLEVASALSIAFNALDDNDGDFQDNKVVVTKTHNSYALADGSH----- 673
Db 931 DHDTERKKLSLL-----IEDTKSESE---KNVIKLNEQIEKL-KGEREKEVRDIQSQL 979
Qy 674 --FIVDPDKLSTIVDLEKILKILIEGLDYSTLRTOTOTOPRDMTAGKALLSKSNLRQGE 731
Db 980 AAKTTDWEKIKITT---LDKVLKEKSDLE-KTNKESVDTLKKEVENLKKEISLLEDQKKDD 1035
Qy 732 KQKFLQ-EAQPFLGRVOLDKAIKAKEKALVTKKATNGOLLERS 774
Db 1036 TTKYKELAAQLTETKSNLSDTMTMELEKTELELKKVRN-ELTEVS 1078

RESULT 27
US-09-583-110-3824
; Sequence 3824, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3824
; LENGTH: 1216
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3824

Query Match 3.5%; Score 175.5; DB 4; Length 1216;
Best Local Similarity 19.4%; Pred. No. 0.0015;
Matches 208; Conservative 133; Mismatches 380; Indels 349; Gaps 46;
Qy 96 LKAVTENFESKQITSSQLEQSKESLSLNKTVFSTSNWEICDPITKG-NTLVGLSKSGV 154
Db 1 MKLIPFLSEEBIQLEQAEANSSKEQ---KKTAE-----QIEAIYTSQNTLVSAAGSG 52
Qy 155 EKLQTDHLVLPQAAADGTQLIQVASFAFTPDKKTAAIAYTSRAGENGESQLDVGKEI 214
Db 53 KTFVNERIL--DQARGVEISQIFSTFTVKAATELKE----- 89
Qy 215 INEGEVFNYSLLKKVTIPTGYKHIGQDAFVDMKNIAEVNLPESLETISDYAPAHALAKOI 274
Db 90 -----RLEKKIS-----KKIQETDDVDLQK-----HLGRQLA 116

/ APPLICANT: Sakowicz, Roman
/ APPLICANT: Goldstein, Lawrence S.B.
/ APPLICANT: Cleveland, Don W.
/ APPLICANT: The Regents of the University of California
/ TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
/ TITLE OF INVENTION: Chromosome Congression
/ FILE REFERENCE: 18557C-000110US
/ CURRENT APPLICATION NUMBER: US/09/150,867
/ CURRENT FILING DATE: 1998-09-10
/ EARLIER APPLICATION NUMBER: US 60/058,645
/ EARLIER FILING DATE: 1997-09-11
/ NUMBER OF SEQ ID NOS: 11
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 2954
/ TYPE: PRT
/ ORGANISM: Xenopus sp.
/ FEATURE:
/ OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
/ OTHER INFORMATION: member of the kinesin superfamily of microtubule
/ OTHER INFORMATION: motor proteins
/ FEATURE:
/ NAME/KEY: DOMAIN
/ LOCATION: (1)..(472)
/ OTHER INFORMATION: kinesin like motor domain
/ FEATURE:
/ NAME/KEY: DOMAIN
/ LOCATION: (473)..(2752)
/ OTHER INFORMATION: rod domain
/ FEATURE:
/ NAME/KEY: DOMAIN
/ LOCATION: (2753)..(2954)
/ OTHER INFORMATION: tail domain
/ US-09-150-867-1

Query Match 3.4%; Score 173.5; DB 4; Length 2954;
Best Local Similarity 20.1%; Pred. No. 0.0076;
Matches 209; Conservative 186; Mismatches 431; Indels 213; Gaps 47;

QY 23 EVFSLVKEPILKOTQASSISGADYAESSGSKLAKINETSQVDPVDTVDFSKRTTPEK 82
DB 1200 EVKLLLEMLLGHLLTDSLS---IEKLELENLEVTETKQLQEBEMKNITIERNELQTN 1255
QY 83 IKONLAKGPBRELKAVNTSESEKQITSGSQLEKSKLSLNTKVPSTSNWEICDPTIK 142
DB 1256 FED--LKAHDSLKQDLSNI--EQSIETQDELRAAQBELREQQLVDSFROQLLD---- 1307
QY 143 GNTLVLSK-----SGVEKLSQTHLVLPQAAQDQTO-LIQVASFAPTPDKKTAIEVT 195
DB 1308 --CSVGISPNHDVANQKVLGEVNSLQSEMLRGERDELQTSKALVSELELRAHVK 1365
QY 196 SPAQNGEITS-QLVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEYN- 253
DB 1366 SVEGENLEITKGLNGLEIKLSE--ESELVLSML-----ENLKEDNN 1407
QY 254 -LPSELETSD-----YAPAHALKQID-LPDNLKA-----TGLAFPD----- 290
DB 1408 KLKQAEYSYSSKENQFSLEEVFSQKLVDEIVLKAQKAAERLEIKORDYFELVQTA 1467
QY 291 --NQITGKLSLRLMLRAERAPKNHIK-TTEPRGN--SLKVIQGEASFQDNDL----- 339
DB 1468 NTNVLVEGKLETPLQ-----ADHEEDSIDRSEEMIKVLGEKLNQYLLERLQE 1517
QY 340 SOLMLPDGLEKTESBAFTGNPGDHYNNRVNLWTSGKNPSGLATENTYV--NPDKSLMQ 397
DB 1518 EKLELSNKLLEILQKEMETS-----VLLKDDLLQOKLESLSLSENIILKENIDITLKH 1567
QY 398 ESPEIDYTKWLEDEFTYQNSVTGSKNG--LQKVKRNKNLEIPQHNGVTITEGDNAF 455
DB 1568 HSDTOAQLOKTOEQLQAKNLAAASDNCPIQEKETSADCVHPLEEKILLTLE---ELH 1624
QY 456 RNVDQFQNTL-RKYDLE--EVKLPSTIRKIGAPAFQS--NNLKSFE--ASDLEEKKEGAFM 510

DB 1625 QKTNEQEKLLHKEKNELEQAQVELKCEVEHLMKSMIESKSSLSLEQHEKCHDTEQQLLALKQ 1684
QY 511 NNRIETLELKOKLVITIGDAAPHINHIYAVLPESVQEIIGRSAPFRONGANNLIIFMSKVKT 570
DB 1685 QMQVVTQEKKELOQTHLTAEDVDH-----LKENI-ELGLN-FKNE-----AQOKT 1728
QY 571 LGEMAFISNRLEHLDLSEKQQLTRIP--VQAFSDNALKEVLLPASLTIREEAFKQNHKL 628
DB 1729 TKEOCLL--NENKELEQSOHRLQCEIBEMLKSLKD---KE-----SALETKESEQKVINLN 1780
QY 629 Q-----LEVASALSHIAFNALDNDGDQFQFNKVVVTKTHN-----SYALADGEHFIV 676
DB 1781 QEMEMVWLEMEELKNSQRTVIAERDQLQDDLRSEVMSIETQDDLRKAQALQOKQKQV- 1839
QY 677 DPKLSSTIVDLKILKIEGLDYSTRQTTQTQTFQDMTTAGKALLSKNSLRQEKQKQF- 735
DB 1840 --QELTSQISVLQEKISLLE-----NOMLYNVATVKTLSERDDLQNSQKHLFS 1886
QY 736 -----LQEAQFPLGRVLDLKAIAKAEKALVTKKATK-NGOLLERSINKAVLAYNNS- 785
DB 1887 BIETLSLSLKEKEFALQAEKADAAKRTIDITEKISNIEBEQLLQQAATNLKETIYERES 1946
QY 786 -----AIKANVVRLEKELDLLTGLVEGKGLAQATMVQGVYLLKTPPLPEYITGL 837
DB 1947 LIQKEQLALNTEHLRETLKSKOLALGHQEQERDEA---ANKVIALTEKMSLEEQINE 2002
QY 838 NVYFDKSG-----KLIIY-----ALDMSDTIGEQQDAYGNPILNVDR-----DNEGVA 881
DB 2003 NVTTLKEGEKETFYLQRPQSQSSQMBELRESLTKDQLQLEBASKEISEATNEIKNL 2062
QY 882 LAVATLADYEGLDIKTLNLSKLSQTSITROVPTAAYHRAGIFQAIQNAARAEQLLPKG 941
DB 2063 TAKTSSLEEEILQNASILNEAVSERNLR-----HSKQQLVSELEQL--SLT 2107
QY 942 THSEKSSSSSASANSKDRGL 960
DB 2108 LKSRDHAPAQSKREKDEAV 2126

RESULT 30
US-09-248-796A-16333
; Sequence 16333, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16333
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16333

Query Match 3.4%; Score 172; DB 4; Length 905;
Best Local Similarity 19.9%; Pred. No. 0.0017;
Matches 178; Conservative 148; Mismatches 328; Indels 242; Gaps 42;

QY 19 THNQEVSFLVKE-----PILKQTOASSISGADYAESSGSKLAKINETSQVDPVDTV 69
DB 41 TNSLSLQKLSAKVLATETPDVSKLTQPSDEEISNNVSTEDGSEK-----ATLPPINNTT 96
QY 70 TDLFSDKRKT-----PEKIKDNLAKGPRBOELKAVNTENTESEKQITSGSQLEQSK 119
DB 97 TDTTDTNGTATTTAAAAAQAQSDLDQL--PAPIKSLKKKFAKYBEKYPILLDAVKIEKK 155


```
Db 1503 QLEKAKKEGAVVNAALSAG--TDKGAGGVAIAVNTVKNFKAEIAGSGSKNKEAGEDKIH 1560
Qy 801 LITGLVEGKGLAQAQTMVQGVYLLKTPLPPEYIYGLNVYFDKSGKLIYALDMSPTIGEG 860
Db 1561 AKHVAVNAKSSVVVNAASGLAISKDAFS---GMSGAWQDLS-----NDTIAKV 1607
Qy 861 QKDAVGNPILVDENE-----GYHALAVATLADYEGLDIKIL-----899
Db 1608 DKGRISADSLNVNANNSILGVNVVAGTIAAGSLSTAVGAAPANNTLNKTSALITGTVKNPF 1667
Qy 900 ---NSKL-----SOLTSIRQVPTAAVHRAGI--FOAIQNAARAEOLLPRGTHSB-K 946
Db 1668 SKNTKVNQALNDSHINVSAGGASIKQAGIGWVNVNRGSDTEALV-----SDSEFE 1723
Qy 947 SSSSESANSKORGLQSNPKNTRGRHSAILPRTGSKGSFYIGLGYTSVALLSLITPAIKK 1006
Db 1724 GVSSFNVDKAD--QKTINTIAGN-----ANGKAAAGVATVAHTNIGKQSVIAIVKNS 1774
Qy 1007 K 1007
Db 1775 K 1775

RESULT 32
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380170
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5178

Query Match 3.3%; Score 169.5; DB 3; Length 1010;
Best Local Similarity 20.5%; Pred. No. 0.003;
Matches 164; Conservative 116; Mismatches 284; Indels 235; Gaps 38;
Qy 19 THNQVFLVLPKQVQASSIS-----GADYAESGSKSKKINETSGPVDV---67
Db 215 TFNDELALYKE--LESSQDKWIEKFPQFNDYGCKILKSPKAKNKITK---ELDDLNH 269
Qy 68 ----TVDLPSDKRTTPEKIK-DNLAKGR-----EOLKAVTENTSEKQITSGSQLEQ 117
Db 270 KYKVNVELSENTKGLKAEIKFDDLKQNYIDKLQKELMIOESKVLITYFRLQSLKK 329
Qy 118 SKESL-----SLNKTVPSTGNWEICDFTKNTGLVGLSKSGVKLSQTDHLVLPQAA 170
Db 330 DKDELVSLEHQSKLNET-----NYH-----NEIKGFQK-----QLEHL-----362
Qy 171 DOTLIQVASF-----AFTPDKKAIAEYTSRAGENGEISQLDVPGKEITINSEGVN 222
Db 363 -STRENEITQFQYLEKQNVFNQDLKISSYQKQPVIEEIKLYSEVNDLITKKEBLT 421
Qy 223 SYLLKKVITYPTGYKHGQDAFVDNKNIAEVNLPESLETISDYAPFAHLAKQIDLDNLKA 282
Db 422 KEMNNKNDFAITIEHYTEIYKXKIIIDESERQKDEKLFDKL-----QLDKSSYLSK 474
Qy 283 IGEAFFQNOIQKGL-SLPLQMLAERAFKSNHIK-----TIEFRGNSLKVIGE---331
Db 475 LKEKKEQLNEBSSITNIDATLIDNDKKDFVNEIKSAMSIGDTFCICGNEIHSIGEID 534
```

```
Qy 332 ---ASFQNDLSQL-----MLPDGLEKIESAFTGNPGDDHYNRVRVLTWYSGKGNPSGLA 383
Db 535 FESIAQNNKIKRLSKKVKIRDBIIKITETRIEELN-----HRENELAFKQEKDISELQ 590
Qy 384 TENTYVNPDKSLMOESPEIDYTKWLEEDFTYQKNSVTG-----FSNKGLOQVKRNKN 435
Db 591 KOLMHLN---QKDEQSSINK---LVENFEKQEKIEVKNIHQFDLDSRRKWTQKEK-----640
Qy 436 LEIPKQHNGVITBIGD---NAFRNV-DFQ-----NKTLRKYDLSEVKLPSTIRKIGA 484
Db 641 LEI-----QINDFERHSQFSSVNDFTYYSHAKQVETVEYENEKTKDLNEL--688
Qy 485 FAFOSNNLKSFEASDDLEIEIKGAPMNRITLBLEKOKLVLTIGDAAAPHINHIYALVLPES 544
Db 689 ---NNKLKTEMNDQKHLTENLTQTSKEINNLEK-----MEKE 723
Qy 545 VOEIGRSAPFRONGANNLIFMGSKVKTLGEMAFLSNRLEHLDLSEKQLTEIPVQAFSD--602
Db 724 MQQLGFESYDO-----VKSAA-----DLSAQKDBIEREINIYNKY 759
Qy 603 -----NALKEVLLPASLKTIRE--BAFKKNHLKQLEVASALSALSHIAFNALDDNDGD---650
Db 760 QSYEIEINRLKELVKGKLLNLESLRQSIETKLNKLDETNSQIATISYK-IDNNSNKNFK 818
Qy 651 -----EOPDNKVVVKTTH-----HNSVALADGEHFIVDDPKLSTIVDLLEKI-----691
Db 819 IKNIQIOLDDELKVKQEIFLLSEILAGKNDYKLT-LENYVL-----IYYLEKLIIPQA 869
Qy 692 ---LKLLEGLDYSTLRQTT 707
Db 870 NQRLSFMGSGNRYQLIRRET 888

RESULT 33
US-09-583-110-3829
; Sequence 3829, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3829
; LENGTH: 1965
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3829

Query Match 3.3%; Score 169.5; DB 4; Length 1965;
Best Local Similarity 19.4%; Pred. No. 0.0079;
Matches 231; Conservative 167; Mismatches 442; Indels 351; Gaps 56;
Qy 4 HLKTVALTTLTVSVVTH-----NOEVFSLVKPEILKQTOASSISGADYA---48
Db 136 HLVLVAGIGNGVLSVVALENHLLNLYNTVDYELTSGEKLPLPKE-----ISGVYIGYI 189
Qy 49 -----ESSGSKLKINET-SGPVDDTVDLPSDKRTTPEKIKD-----N 86
Db 190 KEGKTTSDFEVSNQEKSAATFTKQOKVDYNTVPFVDHPSTVQAIQEQVPSVSTKPTBQ 249
Qy 87 LAKPREQEL-----KAVTENTSEKQITSGSQLESLSLNKTVSTPSNWEICDFTIK 142
Db 250 VVEKPFSTELINPRKEKQSSDSQBLAEHKNLETKKE-----EKISPK-----293
```



```
QY 233 ---TCYKHIG-----QDAFVDMKNIAEYNLPESLETISDYAPAHALAKOI 274
Db 407 QPETGVHKDVSGAIVPAIQPELPEAVVSDKGPEVOPALSKAVITDKGETEV---QP 463
QY 275 DLPDNLKAIGELAFPDNQITGKLSLQRLMLAERAFKSNHIKTIEFRGNSLKVIGESAF 334
Db 464 BSPDT-----VVSCKGPEQVAPLP-----EYKGNIEQVKPETPV 498
QY 335 QNDLSQLMPLDGLKIB-----SEATGNPGDDHNNRVVLTWKSGNPSGLATENT- 387
Db 499 EKT-----EQPEKTEBPVKPTEETPVNNEGTEGTSTI---QEAENPVQAEESTT 549
QY 398 ---VYNPDKS---LWQES-----PEIDYTKLLEEDFTYQKNSVTGFSNKG 426
Db 550 NSEKVPSPDTSSENTGESNPSDSTTSVGESKNKPEHNSKENSEKTEVEVPVN---PNEG 607
QY 427 LQKVRNKNLEIPKQHGVTITEIGDNAFRNV-DPQNK-TLRKYDLBEVKLPSTIRKICA 484
Db 608 TVEGTNSQETEKPVQPAETQNSCKIANENTGEVSNKPSDSKPPVESNQ---EKNGT 664
QY 485 FAFQNNLKSFPASDDLEBEIKGAFMNNRIETLEKDKLVITIGDAAFINHIYAIVLPS 544
Db 665 ATKPNSGNTTSENGQTSPEKKLELN---VSDIELYSQ---TNGT---YRQHVSLDGIPEN 717
QY 545 VQ-----EIGRSAP-----RONGAN---NLIFMGSKVKTIGEMAFSLNLEHLD 585
Db 718 TDTYFVKVKSAPKDVYIPVASITBEKNGQSVYKITAERKLOQELNKYVDFNYFLD 777
QY 586 LSEKQLTEIPQAFSDNALKEVLLPASLKTIRBEAFKKNHLKQLEVASALSHTAFNALD 645
Db 778 KKAKEENTNF---TSFS-NLVK-----AINQN-----PSGYHLA-ASLN 812
QY 646 DNDGEQDNKVVVTHNSYALA---DGEHFIVDPDKLSSTIVDLEKILKLEGLDYSTL 703
Db 813 ANEVELSTDDKSYIKGTFTGQLIGSKGKHY-----AIYNLKK---PLFENLSGATV 861
QY 704 RQTQTQFDMTACKALLSK---SNLROGEKQKFLQEAQFFLGRVYDLQKATAKAEKALVT 761
Db 862 EKLS---LKNVAISGKNIDGSLANBATNGTKIKQVHVGVLAGERGVGGLLAKADQSSIA 918
QY 762 KKATKNGQLLERSIN---KAVLAYN-----NSAIKK-----ANVKLEKELD 800
Db 919 ESSFKG---RIVNTYETDTSYNGLVGLVGLHTGKNSIAKSAKATVITSSNTNRSQIVG 973
QY 801 LITGLVEKGPIQAQTMQGVYLLKTPILPLPYIYIGL-----NVYFDKSGKLIYALDMS 854
Db 974 GLAGLVDRDAQIQDSYAEGDINNKK-----HFGRVAGVAGNLMWDRTSQGVHRAGSLT 1025
QY 855 DTIGEGQKDAYGNPILNVDEDN---EGYHALAVATLADYEGLDIKILNSKLSQTSI--- 909
Db 1026 NVLSD-----VNVNTGNAITGYH-----YTGMMKVANTFSSKANRVFNVTLE 1066
QY 910 ---RQVPTAAYHRAGIFQAIQNAAAEABQLPKPGTHS-----EKSSESSEANGSD 957
Db 1067 KDEVVSKESEFBERG-----TMLDASQIVSKAHSIPLTLPTEVPLSTSGKSDSFSKI 1119
QY 958 RGLQNPKNRGRHSAIPLRTGSKGSFVYGLGYTSVALL---SLITAIKK 1006
Db 1120 AHYQANRALVYKNIKELLPFNKSTIVKYGNLVKNSLLYQKELLSAVMMK 1170
```

RESULT 35

```
US-09-107-532A-3855
; Sequence 3855, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ; ; ; ;
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
```

```
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arianello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3855:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1095 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1095
; SEQUENCE DESCRIPTION: SEQ ID NO: 3855:
US-09-107-532A-3855
```

```
Query Match 3.3%; Score 169; DB 4; Length 1095;
Best Local Similarity 19.4%; Pred. No. 0.0036;
Matches 231; Conservative 171; Mismatches 431; Indels 358; Gaps 62;

QY 8 VALILTVSVVTHNQEV-----FSLVKEPILKQTOASSISGADYABSSG-KSKLKIN 59
Db 46 IAPTITFADEITHPQTVTVELDLAHQYVVEGTFSDGRPMSEVTPHYAVYNGVKQDIFCI 105
QY 60 BTSGPVDVTDLFSDKRTTPEKIDNLAKGRPEQLKAV---TENTESEKQITSGSOLE 116
Db 106 EPGVPIDN-----EFTPGYKPNLPDMPERKALVSLWKAGTDVDTTHVAQKMIW 156
QY 117 QSKESLSLANK-----TVPSTSNWEICDITFK-----GNTLVGLSKSGVE-- 155
Db 157 QEVNGYTLHSIKRLGSAVNIAAEAKINQAIADYQKKPSFHNSTAKIVLGSQTTMTDTN 216
QY 156 ---KLSQTDHLVLPQAAD---GTQLIQVASFATPPD---KKTAAIAEVTSRAGNGEISQ 206
Db 217 NLNLSEFDEVVENTANIDYRVNGNQLI-----ITPNASSKESGVLTKKSAGTGPVAY 270
QY 207 LDVPGKEILNEG-----EVFNSYLLKKV-----TIPTGYKHI-----G 239
Db 271 KKV-GQQTLMAGAIKNTYTVTKIDVTEGSLKIKKVDKESGALVPGTVFHLDFGKNLPA 329
QY 240 QDAFVDMKNIAEYN-LPESLE-TISDYAF-AHLAKQIDLPDNLKAIGELAFPDN---QI 293
Db 330 KDVITDKEGIATLDEIPHGTKVITTEKSVAPYPTIDTTPMTTIKA-GETIYVTSKNAQE 388
QY 294 TGLKSLRPLQRLMLAERAFKSNHIKTIEFRGNSL-----KVIGEASFDNDLSQL-- 342
Db 389 KGQIILDKSGVETGSDLMNDNY-----SLAGNTFAIRKDSPTGIEIVQEMTTDENGHAETPK 444
QY 343 MLPDGLB-----KTESAFTGNPGD-----DHYNNRVVLTWKSGNPSGLAT-ENTY 388
```


445	Db	EIANALELGYTVYTETKASGCFVNTFKPVKVELKYANQTVALTNSVYKQNGQEVTTGL 504
389	Qy	VNPDKSLWQBSPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVKNKNLEIPKQHNGV--- 445
505	Db	TKEDKDTGDKA-----QGVKAVPGTEVTLFTAKDGKAVKWEAFK-PEMWKGTAKAS 554
446	Qy	--TIT-----EIGDNVAFNRVDP-----ONKTLRKYLDEEVKLPISTIRKIGAFAPQSNL 493
555	Db	DEVTTLALDEKNOQAAVKHLAINETYYMQETRAPEGYTLDETQYPVSIKKV----- 603
494	Qy	SFEASDLEELKEGAFMNNRIETLEL--KOKLVITIGDAAFHINHIYAIVLPESVQEIGRS 551
604	Db	-----DNEK-----NAVITRDVTAKQIIRFGDFPKF-----AGSA 636
552	Qy	A-FRQNGANNLIFMGSKVKTILGEMAFSLNRL-----EHLDSLSEOKULTEIPVQAFSDNAL 605
637	Db	AGTAETGFNDLTPKVPSPLEGTNEITGAEDEATTAYNEQLGDFGVGKKEENLP---YGYDLL 693
606	Qy	KEVULLPASLKTITREEAPKQNHILKOLEVASALSHIAFNALDNDGDEOF-----DNKVVVK 660
694	Db	EEVEAP-----EGPQK--ITPLEIRSTPKE--NKEDFPVKSEYVFTITEQDQKQPIK 740
661	Qy	THNSYALADGEHFIQVDPDKLSSTIVLEKILKLEGLDYSLTRQTTQTOFRDMWTYAGKA 720
741	Db	TVTVPYEKLTKNKAFSVSLNRL-----MAYDLPEEEDSLTSLATWKGNGKELTSLDST 792
721	Qy	LIAKSLNLRQGE-KQKFLQBAQFGLGRVDDLKA-----IAKAELKALVTKKATK 766
793	Db	ELVDKLSYNLHEIKEDWYVVAQAI--DVDATKAAQEKDEKAKPVVIAETSATLANKEKTG 850
767	Qy	NGQL-----LERSINKAVLAYN-----NSAIKKANVRLEKELDLTLGL----- 805
851	Db	TWKIQHKLTAEOVLNKTIVLFNYYVYENKEAFEAGDKPVAKDVSLNNQAQTVSCTVEHHVS 910
806	Qy	VEGKGPL--AAQATWVG-----VYLLKTPLELPIPEYIIGL 837
911	Db	IQTKAHLENGSQTFTHGDVDMDDVDSITHDVLGDSKEAFETILYAL-----LPD---GT 962
838	Qy	NVYFDKSGKLIYAL--DMSDTI-----GEGOKDAYGNPILNVDDNEGYHALAV 884
963	Db	NKEIWSGKLEYEVNDKEFTKTVLAKKVDGKYPEGKFTFAE--INYDKD----- 1011
885	Qy	ATLADYEBGLDIKTILNSKLSQTSIROQVPTAAVHRAGIFOAIQNAAEAEQLLPK-PGTH 943
1012	Db	GTINGKHNEDLK-----EKSQTLTPKEVPITL-----STPQKQETP 1047
944	Qy	SEKSSSESSEANSKDRGLQSNPKTNWRGRHSAILPRTGKGSFVYGILGYTSV 994
1048	Db	TVFSGQESS-----LTVKT-----FPQTGEKNVLLIFIGTLI 1082

RESULT 36

```

1  US-09-331-403-2
2  ;
3  ; Sequence 2, Application US/09331403
4  ; Patent No. 6489147
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: ALTMANN-JOHL, Regula; PHILIPPSEN, Peter; ALTHOFER,
9  ;
10 ; Henning; SEULBERGER, Harald.
11 ;
12 ; TITLE OF INVENTION: Adenylyate cyclase gene, and its use
13 ;
14 ; NUMBER OF SEQUENCES: 2
15 ;
16 ; CORRESPONDENCE ADDRESS:
17 ;
18 ; ADDRESSEE: Kell & Weinlauf
19 ; STREET: 1101 Connecticut Avenue
20 ; CITY: Washington
21 ; STATE: D.C.
22 ; COUNTRY: USA
23 ;
24 ; ZIP: 20036
25 ;
26 ; COMPUTER READABLE FORM:
27 ;
28 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
29 ;
30 ; COMPUTER: IBM Ar-compatible, Pentium III processor
31 ;
32 ; OPERATING SYSTEM: Windows 95
33 ;
34 ; SOFTWARE: WordPerfect version 6.1
35 ;
36 ; CURRENT APPLICATION DATA:
37 ;

```

Db 1044 SGEAFKLKLSLR-----VLMNANNLQSLPABISQL-----SQLSVIDVG 1083
Qy 777 KAVLAYN-----NSAIKANVK---RLEKELDLLTGLVEGKPLAQATWQGV 821
Db 1084 SNQKYNTSNYHDWNWQNTLKYLNFSGNTRPEIKSAMGYGTMMHLSDLTVLKQLRVL 1143
Qy 822 YLLKTPPL---PLPEYVIGLVYFDKSGKLIIVALDMSDTIGEGQKDAYGNPIL-----N 871
Db 1144 GLMDVTLNTSRVPD--DGVNFRLRMTGSTMINGEYGVADSLQOTDVSSTRDTPFRFRGN 1201
Qy 872 VDEONEGYHALAVATLADYEGLDIKTLNLSKLSQUTS-----IRQV-----PTAAAYHR 919
Db 1202 EDE-----CLICLYDGINDNTNGHKISQVIRDIYRILVRSVEKYGDSPPAIR 1251
Qy 920 AGI-FOATONAAAEABQLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGHSAILPRT 978
Db 1252 DALREFSLQNLKNEINSSILAAGNNKNSNGNTNNNNNTNNNNNGNNNTNNGESVYLTA 1311
Qy 979 GS-----KGSFVYGILGYTSVALLS 998
Db 1312 DLSGASVTVPVPMKGNITYANIGNATALLS 1342

RESULT 37

US-08-446-137B-2

Sequence 2, Application US/08446137B

Patent No. 6162903

GENERAL INFORMATION:

APPLICANT: Trowern, Angus R.

APPLICANT: Ackinson, Anthony

APPLICANT: Murphy, Jonathan P.

APPLICANT: Laurence, Oliver S.

APPLICANT: Duggleby, Clive J.

TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED

TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSES: SEED and BERRY LLP

STREET: 6100 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,137B

FILING DATE: 22-MAY-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: McMaster, David D.

REGISTRATION NUMBER: 33,963

REFERENCE/DOCKET NUMBER: 100084.406

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1027 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-446-137B-2

Query Match 3.3%; Score 168.5; DB 3; Length 1027;

Best Local Similarity 19.2%; Pred. No. 0.0036;

Matches 221; Conservative 174; Mismatches 436; Indels 323; Gaps 56;

Qy 20 HNOEVSFLVKEPILKQTOASSISGADYAESSGSKLKIINETSQVDDTVTDLFSDKRTT 79

Db 31 HCQNMKINKKLLMAALAGAIUVGGANAYAAERDNTDNNLSMDIEISDAYPDYHGD----- 86
Qy 80 PEKIKDNLAQPRQEL-----KAVTENTE-SEKQITSGSOLEQSKESLSLANKTVPTSN 133
Db 87 ---VSDSV--DPVBEEDIALAKALAEAKETAKKHIDSLNHLSETAKKLAKNDIDSATTI 141
Qy 134 WEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVAFPTDPDKKTAIAE 193
Db 142 NAINDIVARADV---MERKTAEK-EBAEKLAATAETA-----KXHIDE 180
Qy 194 YTSRAGENGEISQDVGKEIINEGEVFNPSYLLKKVTIPTGYKHIGQAFVNDKNIAFVN 253
Db 181 LKHLADTKELAKRIDIDSATTIN---AINDIVAR-----ADVMERKTAKE 223
Qy 254 LPESLETISDYAFAH-----LALKQIDLPDNLKAIGELAFPDNQITGKLS-- 298
Db 224 EAEKLAATAETAKKHIDELKHLADTKELAKRIDIDSATTINDIINVARADVMERKLSK 283
Qy 299 ---LPROLMRL-ABRAPKSNHIKTIEPRGNSLKVIGBASPODNDLSQLMLPDGLEKIBSE 354
Db 284 ETPPEPEEVTIKANLIFADGSTQNAEPKGTAKAVSDA-----YAYADALKXONGE 334
Qy 355 AFTGNPGD-----DHYNNRVVLMTKSGKNPSG-----LATENT 387
Db 335 -YTVDVADKGLTLNFKPAGKKEPEEPKEEVTIKVNLIIFADGKTQTAEFKGTPEEATAKA 393
Qy 388 YVNPDKSLWQESPEIDYTKMLBEDFTYQKNSVT-GPSNKGLOKVRKNKULIPIQHNGVT 446
Db 394 YAYAD-LLAKENGE--YTADLEDG---GNTINIKPAGK-----ETPETPEEPKEEVTIK 441
Qy 447 ITEI-GDNAPRVDPQNKTLRKYLEVKLPSTIRKIGAPAFQSNLKSFEASDDLEBIK 505
Db 442 VNLIIFADGKIQTAEFK-----GTPEEATAKAYAYANLLAKENGEYTDLE 486
Qy 506 EGAPMNNRI-----ETLELKDKLVTIGDAAFHINHIYAIVLVESVQEIQRSAFRQ 555
Db 487 DG---GNTINIKPAGKETPETPEEPKEEVTI---KVNLIIFADGKTQTAEFKG--TPEE 536
Qy 556 NGANNLIPMSKVKTLGEMAFLSNRLE-----HLDLSEQQLTEIPQASDNLAKELV 609
Db 537 ATAERYADLLAKVNGE---YTADLEDGGYTTINIKFAGKEQPGENP-----GITIDSBL 588
Qy 610 LPASLKTIREBAFKKNHLKQLEVASALSASHIAFNALDDNDGDEQFDKNKVVVXTH--HNSYA 667
Db 589 ----LKAKEEAIKE--LKEAGITSDLYFSLINKAKTVEGVEALKNE-ILKAHAGEETPE 641
Qy 668 LADGEHFIVDPDKLSSTIV---DLEKILKLIEGLD---YSTLRQTTQTQFRDMTTAGKAL 721
Db 642 LKDGATVYEBEAAAAAKEALKNDVNNAYEIVQADGRYVYVLK-----I 685
Qy 722 LSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTYKATKNGGOLLERSINKAVLA 781
Db 686 EVADSEEPGEDTPEVQE---GYATYEBEAAAAKEA-----LKEDKVNNAVEV 729
Qy 782 YNNSAIKKANVKRLKELDLTLGLVEGKPLAQATWQGV---YLLKTP-----LPLPE 832
Db 730 VQADGRYVYVYVLKIEDKEDSQPGEPGENP-----GITIDWLLKNKAKDAIKELKE 781
Qy 833 YYIGLVNYFD--KSGKLIYALD-MSDTIGEGQKDAYG-NPILNVDE----- 874
Db 782 AGISSDIYFDAINKAKTVEGVEALKNEILKAHAKPGENPGITIDWLLKNKAAAIKEL 841
Qy 875 -----DNEGYHALAVATL---ADYEG-----LDIKTLNLSKLSQUT 907
Db 842 KEAGITAEYLFNLINKAKTVEGVESLKNELKAHAEKPGENPGITIDWLLKNKAKDAIK 901
Qy 908 SIRO--VPTAAVHRA-----GTFQAIQNAEAEOLLPKPG-----THSEKSSS 949
Db 902 ELKEAGITSDIYFDAINKAKTIEGV-EALKNEILKAHKKEDEPGKPKGEDKKPKPKGE 960
Qy 950 SESANSKDRGLQSNP-----KTNRG-----RHSAILPRTGSKGSFVYGILGYTSVALLSL 999
Db 961 DKKPEDKPKGEDKKPKGKTDKDSPNKKKAKLPAKGEAEIL-----TLAAAL 1013


```

QY 169 AADGTQLIQA--SFAITPD---KKTATAEYTSRAGENCEISQLDVDGKEIINEGEVENS 223
Db 52 AKDGKKDSSAQPQGVAFSDVNTIKRPNPAGTRKXSSNAEVI-----KB-LNKCREENS 104
QY 224 YL--LKKVTIPTGYKHIGQDAPVDNKNIAEAVNLPSLETISDYAFALHAKQI-DLPDN- 279
Db 105 MRLDLSKRSI-----HI-----LPSIKELTQLTLYLYSNKQLSLPAEV 144
QY 280 --LKAIGELAFPDNQTGKLSPLQMLRAERAFKSNHIKTIFFRGNLSLKVIGEASFQDN 337
Db 145 GCLVNLMTLALSNSLT--SLPDSLDNLK-----LRMLDLRHNKLRLEIPSVVYRLD 194
QY 338 DLSQMLPDGLEKIESEAFTPGDDHYNNRVLMWKSGKQPSGLATENTYVNPDKSLWQ 397
Db 195 SLTTLYL-----RFRNRTTVEKDKIKLSLMSLSIRENKIKQLPA 234
QY 398 ESPEIDYTKWLEEDFTYQKNSVTGFSNKGQLQVKNKLEIPKQHGVTITEIGD-NAPR 456
Db 235 EIGEL-----CNLIT-----LDVAHQLEHLPK-----EIGNCTQIT 266
QY 457 NVDFQNKTKLYKDYLEVKLPSTIRKIGAPA-----FOSNNL-----KSFASDDLLEIKEGAP 509
Db 267 NLDLQHNL-----LDLPDTIGNLSLSIRGLRYNRLSATPRSLAKCSALEELN---L 316
QY 510 MNNRIETL--ELKDKLVITIGDAAPHIN--HIYAIVLPESVQIEGRSAFRQNGANNLIF-M 564
Db 317 ENNNISTIPESLLSLVXLNSLTLARNCFQLYPVGGPSQFSTIYSLNMEHNRINKIPFGI 376
QY 565 GSKVKTGLGEMAFSLNLEHLDLS-----EQQLTEIP-----VOAFSD 602
Db 377 FBRKVLKLNKMDKQLSLPLDFGTWTSWVNLATNQLTKIPEDVSLVSLVLEVLILSN 436
QY 603 NALKVELLP---ASLKTRIBEEAFKXHLKQL--EVA 633
Db 437 NLLKK--LPHGLGNLRKLRELDLEENKLESPLNEIA 470

```

RESULT 40

```

US-08-053-614-2
; Sequence 2, Application US/08053614
; Patent No. 5403924
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: TUMURU, MURALI K.R.
; TITLE OF INVENTION: THE taga GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSES: NEEDLE & ROSENBERG, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/053,614
; FILING DATE: 19930426
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 859 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-053-614-2

Query Match      3.3%; Score 167.5; DB 1; Length 859;
Best Local Similarity 20.4%; Pred. No. 0.0032;
Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;

QY 27 LVKPEILKQTQASSISGADYAESGSKLKNINETSQVDDVTVDLPSDKRTTPEKIKDN 86
Db 1 MTNETIQDPQTEAFAFPQPFNNLQVAFK-----VDNAVASYDPDQKPIVDK---- 49
QY 87 LAKGPREQELKAVTENTESEKQITSGSQLESLSLNTKV--PSTSNWBEICDPIITKGN 144
Db 50 -----NDRDNRQAFEG--ISQLEREYS-NKAIKNPTKQNFQSFDFINKSN 91
QY 145 TLVG-----LSKSGVEKLSQTDHLVLPQAADGTQLIQVASFAP-----TP 185
Db 92 DLINKDNLIVVESSTKSFQKPGDQRYRIFTSWVSHQNDPSKINTRCIRNFMEHTIQPIP 151
QY 186 DKTAIAEYTSRAGEN--GEI--SOLDVDGKEIINEGEVENSYLLKKVTI-----PTGYK 236
Db 152 DDKEK-AEPLKSAKQSFAGIIIGNQIRTDQKFM---GVFDESLEKQEAKEKNGPTGSD 206
QY 237 HIGQDAFY---DNKNIAEAVNLPSLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQ 292
Db 207 WL--DIFLSFIDFKKQSSDVK-----EAINQEPLPHVQ-----PD-----IATSTTH 246
QY 293 ITGKLSLPROQLMRLAERA---PKSNHIKTIETPRG-----NSLKVIGEASFQDND 338
Db 247 IOGLPPESRDL--LDERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLI-----HNNT 298
QY 339 LSQMLM--PDGLEKIE--SEAFITGNGP-----DDHYNNRVLMWKSGKQPSG 381
Db 299 LSSVLMSGHSDGIEPEKVSLLYAGNGFGCAKHDWATVGYKQQGNNVATIINVHMKSGS 358
QY 382 LAT---ENTVYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGQLQVKNK--NL 436
Db 359 LVIAGEGKGINNP-----SPLYKEDQLTGSQALSQSEIQNKIDFM 400
QY 437 EIPKHQNGV--TITEIGNAPRN--VDPOKNTLRYKDYLEVKLPSTIRKIG--AFAPQSN 490
Db 401 EFLAQNNAKLDLSLSEKEKEFRNEIKDFQ-KDSKPY-----LDALGNDRIFAVSK 449
QY 491 NLKSFASDDLLEIKEG-----AFNNRITETLEKDKLVITIGDAAPHINHIYAILVPE 543
Db 450 --KDPKHSALITEFNGKGLDLSYTLKVMGKQIKALD-REKNTVLOGNLKH-----D 496
QY 544 SVQBIGRSAPRQNGANNLIFMGSKVKTGLGEMAFSLNRLHLDLSEK--Q 591
Db 497 GVMFVNTSNPKYTNASK-----SPNKGVG---VTNGVSHLEAGFSKAVFNLPLNLA 547
QY 592 LTEIPVQAFPSNALKEVLLPASLKTIRBEAFKXHLKQLEVASALSAPALDNDGDGE 651
Db 548 ITSVRRDLEDKLIAGLSPOEANKLVKDFLSSN--KEL-VGKALNF----- 591
QY 652 QFDNKVVVKTTHNSYALADGHEHFIVDPDKLSSTTVDLEKILKLEGLDYSLRQTQTQF 711
Db 592 ---NKAVAEAKVTG-----NYDEVKRAQDKLEKSLKKEHLE----- 625
QY 712 RDMTTAGKALLSKNLROGEKQKFLQBAQFFLGRVLDLDAJAKAEK----ALVTYKATPN 767
Db 626 KDV---AKNLESKS---GNKNKMEAKAQ-----ANSQKOEIFALINKEANRD 666
QY 768 GQLLE-----RSINKAVLAYNNSAI-----KKNVKELEKELDLTLGLV 806
Db 667 ARAIYAYQNLKGIKRELSDKLENINKOLKDYFSKDFGFKNGKNKDFSKAEETLKALKGSV 726
QY 807 EGKGPLAQATWVGQVYLLKTP---LPLPEYVIGLVNVPDKSGKLIYALDMDSDTIGEGQKD 863
Db 727 KDLG-----INPEWISKVENLNAALNEFKNGKNKDFSKVTQA-----KSD-----QEN 769

```


STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,306
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-034-306-2

Query Match 3.3%; Score 167.5; DB 2; Length 859;
Best Local Similarity 20.4%; Pred. No. 0.0032;
Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;

QY 27 LVKEPILKQTAASSISGADYAESGSKLKNETSGPVDVTDLFSDKRTTPKIKDN 86
DB 1 MTNETIDQPTAEAFNPQPINNIVQAFLEK-----VDNAVSYPDPQPIVDK---- 49
QY 87 LAKGPRBELKAVTENTSEKQITGSQLQESLSLKNVT--PSTNWEICDITKGN 144
DB 50 -----NDRDNROAFEG--ISQUREEYS-NKAIKNPTKKNQYFSDPINKSN 91
QY 145 TLVG-----LSKSGVEKLQTHLVLPSQAADGTQLOIVASPAF-----TP 185
DB 92 DLINKNLIVSESSYKSFQKQDQYRIFTSWSHQNDPSKINRCIRNFMEHTTQPIIP 151
QY 186 DKTAIAEYTSRAGEN--GEI--SQLVDGKEIINEGEVFNLSYLLKQVTI-----PTGYK 236
DB 152 DDKEK-AEFLKSAKQSFAGIIGNQIRTDQKPM---GVFDSLSKERQEAENKGGTGD 206
QY 237 HIGQDAFV---DNKNIAEVLNPESELETISDYAFAPHAHLKQIDLPDNLKAIGELAFPNQ 292
DB 207 WL--DIFLSFIDFKQSSDVK-----EAINQBPFLPHVQ-----IATSTH 246
QY 293 ITGKLSLPRQLMLAERA-----FKSNHKTIFERG-----NSLKVGEASFQND 338
DB 247 IGLPPEERDL--LDERGNFSKFTTGDMEMLDVEGADMDPNYKFNQLLI-----HNT 298
QY 339 LSQML--PDGLEKIE--SEAFGNPQ-----DDHYNRVVLMTKSGKNPSG 381
DB 299 LSSVLGSHDGLIEPEKVSLLYAGNGFGAKHDWNAVGVYKQDQGNVATIINVHMGSG 358
QY 382 LAT---ENTYVNPDKSLWQESPEIDYTKWLEBDFYQKNSVTGFSNKGQKVRNK--NL 436
DB 359 LVIAAGEKGINP-----SFYLYKEDQLTGSQRALSQEBIQNKIDPM 400
QY 437 EIPKQHNGV--TITIGDNAPFN--VDFQNTLRYKVDLEEVKLPTIRKIG--AFAPQSN 490
DB 401 EFLAQNAKLDSLSKEKEKFNETKQF-KDSKPY-----LDALGNDRITAFVSK 449
QY 491 NLKSPFASDLEIEKEG-----AFMNNRIETLEBKDKLVITGDAAFHINHIYAILVPE 543
DB 450 --KDPKHSALITFENKGDLSYTLKVMGKKQIKALD-REKNVTLOGNLKH-----D 496

QY 544 SVQEIGBSAFQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSEQK-----Q 591
DB 497 GVMFVNTYSNFKYTNASK-----SPNKGVG-----VTNGVSHLEAGFSKAVFNLNLA 547
QY 592 LTEIPVQAFSDNALKEVLLPASLKTIREEAPKKNHLKQLEVASALSALSHIAFNALDNDGDE 651
DB 548 ITSVVRLDEKLIAGLSPOEANKLVKDFLSSN--KEL-VGKALNP----- 591
QY 652 QFDMKVVVTHNSYALADGEHFIVDPDKLSSTIVDLKLEKILKLEGLDYSTLRTTOTQF 711
DB 592 ---NKAVAEAKNTG-----NYDEVKRAQKOLEKSLKREHLE----- 625
QY 712 RDMTTAGKALLSKSNLRQGEKQKFLQEAQPLGRVLDLDAKAKAEK---ALVTKKATKN 767
DB 626 KDV---AKNLESKS-----GNQNMBAQAQ-----ANSQKQEIFALINKEARNR 666
QY 768 QOLLE-----RSINKAVLAYNNSAI-----KCANVVRLEKELDLTLGLV 806
DB 667 ARAIYAAQLKGIKRELSDKLENINKDKLQPSKSGDFGKNGKNDKPSKAEETLAKLKGSV 726
QY 807 EGKGLAQAQWVGYYLLKTP---LPIPEYYIGLVNVPDKSGKLIYALDMSDTTIGEGKD 863
DB 727 KDLG-----INPEWISKVENLNAALNEPKNGKNDKPSKVTQA---KSD-----QEN 769
QY 864 AYGNPILN-----VDEDNQGYHALAVATLA-DYEGLE-----DIKTILNSKLSQTSIRQ 911
DB 770 SIKDVIINQKITDKVDELNQ---AVSVAKTACDPSGVEQALADLKNFSKEQLAQ--QAQK 824
QY 912 VPTAAYHRAGIFQAIQN 928
DB 825 NESFVNGKSEIYQSVKN 841

RESULT 43
US-09-259-437-2
; Sequence 2, Application US/09259437
; Patent No. 6153390
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/259,437
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:
 LENGTH: 859 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

Query Match 3.3%; Score 167.5; DB 3; Length 859;
Best Local Similarity 20.4%; Pred. No. 0.0032;
Matches 212; Conservative 150; Mismatches 344; Indels 331

Qy	27	LVKEPILKQTCASSISGADYAESGSKLKINETSGPVDVTDLFSDKTTEPEKIKDN	86
Db	1	MTNETIDQOQTEAAFPNQPINLQVAFLK-----VDNAVASYDDQKPIVDK----	49
Qy	87	LAKGPREQLKAVTENTESEKQITTSQSLEQSKESLSLANKTV--PSTSNWEICDPITKGN	144
Db	50	-----NDRDNQATEG--ISUREEYS--NKAIKMPTKKQNQVPSDFPINKSN	91
Qy	145	TLVG-----LSKSGVEKLSQTDHLVLPQQAADGTOLIOVASFAF-----TP	185
Db	92	DLINKDNLI VVESSTKSFQKQDQRYRIFTSWVSHQNDPSKINTFCINFMHEHTIQPIP	151
Qy	186	DKKTAIAEYTSRAGEN--GEI--SOLDVDGKEIINEGEVFNYSLLKKVTI-----P	236
Db	152	DDKEK-AEFLKSAKQSPAGIILIGNQIRTDQKFM---GVFDES LKERQEA EKNGSGPTG	206
Qy	237	HIGQDAFV---DNKNIAEVLNPSLETSIDYAFAHALKQIDPDNLKALGELAFPDNQ	292
Db	207	WL--DIFLSFTFDKQSSDVK-----EAINOEPLPHVQ-----PD-----IAT	246
Qy	293	ITGKLSLPSQLMRLAERA-----FKSNHIKTIETFRG-----NSLKVIGEASFO	338
Db	247	IQGLUPPESRDJ--JDERGNFSKFTLGDMEMLDVEGVAMDPNYKFNQLLI-----	298
Qy	339	LSQMLM--PDGLEKIE--SEAFGTGNPG-----DDHYNNRNVVLWTSGKGNPSG	381
Db	299	LSSVLMGSHDGIIEPEKVSLLYAGNGGFGAKHDWNAVGYKQOQGNVATIINVMKNG	358
Qy	382	LAT---ENTVYNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGSPNKGLOKVRNK--	436
Db	359	LVIAGGEKGINP-----SFVLYKEDQLTGSORALSQSEIQNKIDFM	400
Qy	437	EIPKQHNQV--TITEIGDNAFRN--VDFONKTLRYDLEEVKLPSTIRKIG--AFAP	490
Db	401	EFLAQNAKLDSLSEKEKEPRNEIKFQ--KDSKPY-----LDALGNDRIFAYSK	449
Qy	491	NKLSFEASDDLEEIKEG-----AFMNNRIETLELKDCLVTIGDAAAPHINHIAI	543
Db	450	--KDPKHSALITFNKGDLSVTLKVMGKKQIKALD-REKNVTLQGNLKH-----	496
Qy	544	SVQBIGRSAFQNGANNLIFMGSKVKTLIGEMAFLSNRLEHLDLSQK-----Q	591
Db	497	GVMFVNTSNFYKNASK-----SPNKGVG-----VTNGVSHLEAGSKVAVFNL	547
Qy	592	LTBITPQVAFSNAKVELLPSLTKTIRIEAPKKNHLKQLEVASALSHIAFNALDNDG	651
Db	548	ITSVVRDLEDKLIAKGLSPQEAANKLVKDFLSSN--KEL--VGKALNF-----	591
Qy	652	QFDKNVVKTHNSYALADGHEFIVDPDKLSTIVDLEKILKLEGLDYSLRTQTTQTF	711
Db	592	---NKVAEAKNTG-----NYDEVKRAQKLESLKREHLE-----	625
Qy	712	RDMTTAGKALLSKNLROGEKQFLQEAQFFLGRVLDLQATAKEK---ALVTKKATKN	767
Db	626	KDV---AKNLESKS-----GNKNKMEAKAQ-----ANSOKEIFALINKEANRD	666
Qy	768	GOLLE-----RSINKAVLAYNNSAI-----KKANVKREKEKELDLLTGLV	806
Db	667	ARATAYAQNKGIKRELSDKLENINKOLKDFSKSFDGFKNGKNKDFSKAETLTKALGS	726
Qy	807	EGKGPLAQATWQGVYLLKTP---LPLPEYVIGLNVVYFDRSGKLIYALDMSDTTIG	863

Db	727	KDLG-----INPWIISKVENLNAALNEFKNGKVKQFQKVQQA-----KSD-----QEN	769
Qy	864	AVNGPILN-----VDENEGYHALAVATLA-DVEGL-----DIKTILNKLQSLTSIRQ	911
Db	770	STKDIIIIKQITDKVDLNLQ-----AVSVAKIACDFSGVEQALADLNKFSKQLAQ--QAQK	824
Qy	912	VPTAAAYHRAGIFAQION	928
Db	825	NEGFNNGKSEIYQSVKN	841

RESULT 44

REG0011 14
 PCT-US93-09782-2
 Sequence 2, Application PC/TUS9309782
 GENERAL INFORMATION:
 APPLICANT: COVER, TIMOTHY L.
 APPLICANT: BLASER, MARTIN J.
 APPLICANT: TUMMURU, MURALI K.R.
 TITLE OF INVENTION: THE taga GENE AND METHODS FOR DETECTING
 TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 STREET: 133 Carnegie Way, Suite 400
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: USA
 ZIP: 30303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/09782
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: SPRATT, GWENDOLYN D.
 REGISTRATION NUMBER: 36,016
 REFERENCE/DOCKET NUMBER: 2200.009
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 859 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US93-09782-2

Query Match 3.3%; Score 167.5; DB 5; Length 859;
Best Local Similarity 20.4%; Pred. No. 0.0032;
Matches 212; Conservative 150; Mismatches 344; Indels 331

Qy	27	LKVEPILKQTOASSISGADYABSSGSKL	KINETSGPVD	DTVTDLPSDKRRTPEKIKON	86		
Db	1	MTNETIQOQPTAEAFNPQOFINLQVAF	LK-----VDNA	VADYDPDQDKPVDK----	49		
Qy	87	LAKGPREQELKAVTENTSEKQITSGS	QLEQSKESLS	LNKTV--PSTSNWEICDFTIKG	144		
Db	50	-----NDRDNQOAFEG--ISQ	REYS-NKAI	KNPTKQNYFSDPINKSN	91		
Qy	145	TLVG-----LSKSGVEKL	SQTDHLVLP	QSAADGTLQVASFAP-----TP	185		
Db	92	DLINKNLIVBSSSTKSFQK	QDQRYRIFT	SWSHQNDPSKINTRCIRNFWEHTIOPP	151		
Qy	186	DKKTAIAEYTSAGEN--GEI--	SOLDVDG	EELINEGEVENSYLLKKVTI-----PTGYK	236		
Db	152	DKKEK-AEPLKSAKOSFAGI	IIGHQIRTDQ	KFM---GVFDES	LKEHQEAKNGGGTGGD	206	
Qy	237	HIGQDAFV-----DNKNIAE	VNLP	PSLETSIDSYAF	LAHLALQIDULP	NLKAIGELAFDQ	292


```

Db      825 NESFNVGKSEIYQSVKN 841

RESULT 47
US-09-034-306-4
; Sequence 4, Application US/09034306
; Patent No. 5876943
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,306
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-034-306-4

Query Match      3.3%; Score 167.5; DB 2; Length 1181;
Best Local Similarity 20.4%; Pred. No. 0.0052;
Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;

Qy      27 LVKPEILQTOASSISGADYAESGKGLKINETSGPVDVTDLPFSDKRTTPEKIDN 86
Db      1 MTNETIDQPTQTEAFNQQFINNLQVAFK-----VDNAVSYDDQKPIVDK---- 49
Qy      87 LAKGPREOELKAVTENTSEKQITSGSQLESKESLAKTV--PSTSNWEICDPIFGN 144
Db      50 -----NDRDNRQAPEG--ISQUREEYS-NKAINKPTKKNQVFSDFINKSN 91
Qy      145 TLVG-----LSKSGVEKLSQTHLVLPQAAAGTQLIQVASPAF-----TP 185
Db      92 DLINKNDLIVVESSTKSPKFGDQRYRIFTSWVSHQNDPFSKINTRCIRNFMEHTIQPIP 151
Qy      186 DKTKAIAEYTSRAGEN--GEI--SOLDVDGKEIINEGEVFNLSYLLKKVTI-----PTGYK 236
Db      152 DDKEK-AEFLKSAKQSPAGIIGNIQTIDQFM-----GVFDESLKERQAEKNGGPTGCD 206
Qy      237 HIGQDAFV----DNKNIAEVLNPELETISDYFAHLAKQIDLPDNLKAIGELAFPNQ 292
Db      207 WL--DIFLSFIPDKKQSSDVK-----EAINQEPPLPHVQ-----PD-----IATSTTH 246

293 ITGKLSLPROLMRLAERA-----FKSNHIKTIETPRG-----NSLKVTIGEASPDND 338
247 IQGLPPPSRDL--LDERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLI-----HNNT 298
339 LSQMLL--PDGLEKIE--SEAFITGNPG-----DDHYNNRVVLWTKSGKNPSG 381
299 LSSVLGSHDGIPEKVSLLYAGNGGFGCAKHDWNAITVGYKQOQGNVATIIINVEHMGSG 358
382 LAT---ENTYVNPDKSLWQESPBIIDYTKWLEEDFTYQNSVTSVGSNGLQVKVKNK--NL 436
359 LVIAGEGKGINNP-----SFYLYKEDQLTGSQALSQBEIQNKIDFM 400
437 EIPQHNQV--TITEIGNAPRN--VDFONKTLRKDYDLEEVKLPSTIRKIG--AFAPQSN 490
401 EFLAQNAKLDSLSEKSEKFRNEIKDFQ--KDSKPY-----LDALGNDRIAPVSK 449
491 NLKSFASDDLEIEKEG-----AFMNNRITETLELKDCLVTIGDAAPHINHIYAIVLPE 543
450 --KDPKHSALITEFNKGDLSYTLKVMGKKQIKALD--REKNVTLOGNLKH-----D 496
544 SVQEIGRSAPRQNGANNLIPMGSKVTLGEMAFNLSNLEHLDLSEK-----Q 591
497 GVMFVNTSNPKYTNASK-----SPNKGVG---VTNGVSHLEAGFSKVAVFNLNLA 547
592 LTEIPVQAFSDNALKEVLLPASLKTIREEAPKKNHLKQLEVASALSHTAFNALDDNDGDE 651
548 ITSVVRRDLEDKLIAGLSPOEANKVJDFLSSN--KEL--VGKALNF----- 591
652 QFDNKVVVKTTHNSYALADGHEHFTVDPDKLSSTTVDLKILKLEGLDYSLRTQTOTQF 711
592 ---NKAVAEAKVTG-----NYDEVKBAQKLEKSLKKEHLE----- 625
712 RDMTTAGKALLSKNLRQGEKQKFLQEAQFPLGRVDLDKATAKAEK-----ALVTYKATKN 767
626 KDV---AKNLESKS-----GNKNKMEAKQ-----ANSQKQEIFALINKENRD 666
768 GQLE-----RSINKAVLAVNNSAI-----KXANVRKLEKELDLTLGLV 806
667 ARATAYANLKGIRKELSDKLENINKOLKDFSKDFGKNGKNDKDFSKAEEETLKALGSV 726
807 EGKGLAQATWVGQVYLLKTP---LPLPEYVIGLVNFDKSGKLIYALDMSDTTIGEGQKD 863
727 KDLG-----INPEWISKVENLNAALNEFPNGKNGKDFSKVTOA-----KSD-----QEN 769
864 AYGNPILN-----VDENEGYHALAVATLA-DYEGL-----DIKTILNSKLSQTSIRQ 911
770 SIKDVIINQKITDKVDLNLQ---AVSVAKIACDFSGVEQALADLKNFSKEQLAQ--QAQK 824
912 VPTAAVHRAGIFQAION 928
825 NESFNVGKSEIYQSVKN 841

RESULT 48
US-09-259-437-4
; Sequence 4, Application US/09259437
; Patent No. 6153390
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/259,437
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-259-437-4

Query Match 3.3%; Score 167.5; DB 3; Length 1181;
Best Local Similarity 20.4%; Pred. No. 0.0052;
Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;

QY 27 LVKPEILKQTOASSISGADYAESGSKLKINETSGPVDVTDVDFSKRTTPEKIKON 86
DB 1 MTNETIDQPOTEAFNPQOFINNQLVAFLK-----VDNAVASYPDQKPIVDK---- 49
QY 87 LAKPREQELKAVTESEKQITSGSQLEBSKESLSLNTV--PSTSWWEICDFITKON 144
DB 50 -----NDRNDRQAFEG--ISQLEEYS-NKAIKPTKQYFSDFIKSN 91
QY 145 TLVG-----LSKSGVEKLSQTDHLVLPQQAADGTQLIOVASPAP-----TP 185
DB 92 DLINKDNLIVSESTKSFQKGDQRYRFTSWSHQNDPSKINTRCIRNMEHTIOPPIP 151
QY 186 DKRTAIAYTSRAGEN--GEI--SQLDVPGKEIINEGFVNSYLLKKVTI-----PTGYK 236
DB 152 DDKEK-AEFLKSAKQFAGIIGNQIRTDQKFM---GVFDESLEKQEAERKNGGPTGD 206
QY 237 HIGQDAFV---DNKIAEVLNPLESETISDFAHALKQDIDPNLKAIGELAPFDNQ 292
DB 207 WL--DIFLSFIDFKQSSDVK-----EAINQEPLPHVQ-----PD-----IATSTTH 246
QY 293 ITGKLSLPLQMLAEARA---FKSNHIKTIIEPRG-----NSLKVIGEASFQDND 338
DB 247 IQLPPESDL--LDERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLI-----HNNT 298
QY 339 LSQJML--PDGLEKIE--SEAFNGP-----DHNINRVVLWTKSGKNPSG 381
DB 299 LSSVLGSHDGIPEKVSLLYAGNGGFGAKHDNATVGYKQDQGNVATIIINVHMGSG 358
QY 382 LAT---ENTVVPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRNK--NL 436
DB 359 LVIAEGKEGINNP-----SPLYKEDQLTGSQALSQEEIQNKIDPM 400
QY 437 EIPKQNGV--TITEIGNAFRN--VDQNKTLRKVDLEVKLPSTIRKIG--AFAPQSN 490
DB 401 EFLAQNNAKLDLSSEKEKEKFRNEIKDFQ--KOSKPY-----LDALGNDRIFAJSK 449
QY 491 NLKSPASDDELEIKEG-----AFWNNRIETLELKDVLVTIGDAAFHINHIYAVLPE 543
DB 450 --KDPKHSALITFENKGDLSYTLKVMGKKQIKALD--REKNVTLOGLKH-----D 496
QY 544 SVOEIGRQAFRQNGANNLIFMGSKVTKLGEAFPLSNRLEHLDLSEOK-----Q 591
DB 497 GVMFVNYSNFKYTNASK-----SPNKGVG-----VTNGVSHLEAGFSKAVFNLNLLNLA 547

QY 592 LTEIPVQAFSDNALKEVLLPASLKTIRBEAFKKNHLKOLEVASALSHIAFNALDDNDGDE 651
DB 548 ITSUVRRDLEDKLIAGLSPOEANKLVKDFLSSN--KEL-VGKALNF-----591
QY 652 QPDNKVVVTHNSVALADGEHFIVDPDKLSSTIVDLEKILKILIEGLDYSTLRQTTQTOF 711
DB 592 ---NKAVAEAKNTG-----NYDEVKRAQDKLEKSLKKREHLE-----625
QY 712 RDMTTAGKALLSKSNLROGEKOKFLOEAQFFLGRVDLDKAIKAKEK---ALVTKKATKN 767
DB 626 KDV---AKNLESK---GNKNMEAKAQ-----ANSQKDEIFALINKKAND 666
QY 768 GOLLE-----RSINKAVLAYNNSAI-----KCANVVRLEKELDLITGLV 806
DB 667 ARAIYAQNLIKGIKRELSKLENINKDKPFSKPDGPKNGKNDKFSKAEETLKALKGVS 726
QY 807 EKGGLPQAQTMVQGVYLLKTP---LPLPEYIYIGLVYFDKSGKLIYALDMSDTIGEGQKD 863
DB 727 KDLG-----INPEWISKVENLNAALNEFKNGKNDKFSKVTQA-----KSD---QEN 769
QY 864 AYGNPILN-----VDEDNVGHALAVATLA-DYEGE-----DIKTILNSKLSOLTSIRQ 911
DB 770 SIKOVIIINQIKTDKVDLNLQ---AVSVAKIACDFSGVBOALADLNKFSKBQLAQ--QAK 824
QY 912 VPTAAVHRAGIFQAION 928
DB 825 NESFNVGKSEIYQSVKN 841

RESULT 49
PCT-US93-09782-4
Sequence 4. Application PC/TUS9309782
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TUMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSES: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-09782-4

Query Match 3.3%; Score 167.5; DB 5; Length 1181;
Best Local Similarity 20.4%; Pred. No. 0.0052;

136 YQNLVFNMDYQDLNLSGGHRRHRETDYDT-----EKWFEISH----- 174
171 DGTQLIQVASFAPTPDKKTAIAEVTSPAGENGESQLDVGDKIINEGE--VFN5YLLKK 228
175 --DQKNVSIYA---NQKTSYCWMLKDYFNKNYDHLNVSINRLEAEAFYAFDDFSQTI 229
229 VTPTGYKHIGODAFVDNK-----NIAEV-----NLPESLET 260
230 KLTNNSYQTVNIDVFNDFNLCILALRFLLSERFNLIRSSYTRNQYNFEKIGELLET 289
261 ISDYAFALHAKQIDLPDNLKAGELAFPDNQITKGLS----- 298
290 IFAVFVSHRHLQIHLQVPCFAFYLVNSSQISVKDSQSQVYSFSTDLKLVDTNKVQDY 349
299 -----LPROLMRLAERAPKSNHKTIEPRGNSIKVIGEAS-----FQNDL 339
350 FKFLQEFPR-LTHVSOQAIPVSNATNAVENLVLKVKVGHANLNLVSIPTQFNDFYFVNL 408
340 SQMLPFDGLE-----KIBSEAFPGDGHYNNRVVLWTKSGKNPGLATE-NTYVNP 391
409 QHLKLEFLEPNILTKQLEN-----LLLSIKOSKNLKFRLNFYTYVAQ 453
392 DKSLWQSPIDYTKLMBEDFTQKNSVTGFSNKGQKVRNKNLEIPKOHNGVTTIEIG 451
454 ETSRKO-----ILKOATT-----IKNLKNKNQEBETPE---TKDET 487
452 DNAFRNVDFONKTRKYDLEE--VKLPST---IRKIGAFQFQSNLKSFEAS--DDLEE 504
488 SETSGMKFDFHUSELSETEDFSNLQATQIEIVDSLHLKLLIRSTNLKFKLSYKEMEKS 547
505 KEGAF--MNNRIETL-ELKDKLVITIGDAAAPHINHIYAVLPESV-----QEI--GR 550
548 KMDTFIDLKNIYETLNNLKCSVNI--HGNISVELTNKDSFTFYKFKLTNLQELQHA 605
551 SAFRQGANLIPGSKVTKLGEMAFSLNRLEHLDSEQKQLTEIPVQAFSDNALKEVLL 610
606 YTFKQNE-----POFNVK-----SAKIESSLESLEDIDSLCKSIASCNKLQVNI 652
611 PASL---KTIREAFKQNLH---KOLEVASALSHIAFNALDDNDGDEQFNDKVVVKTTHN 664
653 IASLLYPNNIQKPNFKPNLLFFQFQELKNLNVINCILD----- 694
665 SYALADGEHFIVDPDKLSTIVLEKILKLEGLDYSTLROTTO-----TQPRD 713
695 QHILNISELEKKNKIKAFI--LKRYLLOYLDYTKLFTLQQLPELNOVINOOLEE 752
714 MTAGKALLSKSNLROCKQK-----FLOBAQFFLGRVLDLKAIA---KAEKALVTK 762
753 LTVSE-----VHKQVWENHKQAFYELCEPIKESSTQLQIDFDQNTVSDSIIKILESI 808
763 KATNGQLLERSINKAVLAYNNSAIKANVRLEKELDLTLGLVEGKGLQAQNTWQGY 822
809 SESKHYHLRNPQ-----SSLSIKSEN-----EBIQELLKACDEKGLVKA----- 851
823 LKLTPLPEYIYGLNVYFD 842
852 YKFPCLCP-----TGTYD 866

RESULT 51
US-08-851-843A-54
; Sequence 54, Application US/08851843A
; Patent No. 6093809
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6093809el Telomerase

NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-851-843A-54

Query Match 3.3%; Score 167; DB 3; Length 872;
Best Local Similarity 20.5%; Pred. No. 0.0036;
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;

Qy 13 TTVSVVTHNQVFP-----SLVKEPILK-----QTOASSSISGADYAESGSKLK 57
Db 17 TNLDFVLQNLVYKSGIEHYKTOQOIKEEDLKLKFNQDQDGNNGDDDEENNNSN-K 75
Qy 58 INETSGPVDDTVDLFSDKRTTPEIKD-----NLAKGPREQBELKAVTENTESEKQIT 110
Db 76 QQELLRVNOIKQVQLIKKVGSKVEKDLNLNEDENKKNGLSEQQVKEBQLATITEEQVK 135
Qy 111 SCSQLEQSKESLSLNTKVPSTNSWEICDPTIKGNTLVGLSKSGVKEKLSQTHLVLPSQAA 170
Db 136 YQNLVFNMDYQDLNLSGGHRRHRETDYDT-----EKWFEISH----- 174
Qy 171 DGTQLIQVASFAPTPDKKTAIAEVTSPAGENGESQLDVGDKIINEGE--VFN5YLLKK 228
Db 175 --DQKNVSIYA---NQKTSYCWMLKDYFNKNYDHLNVSINRLEAEAFYAFDDFSQTI 229
Qy 229 VTPTGYKHIGODAFVDNK-----NIAEV-----NLPESLET 260
Db 230 KLTNNSYQTVNIDVFNDFNLCILALRFLLSERFNLIRSSYTRNQYNFEKIGELLET 289
Qy 261 ISDYAFALHAKQIDLPDNLKAGELAFPDNQITKGLS----- 298
Db 290 IFAVFVSHRHLQIHLQVPCFAFYLVNSSQISVKDSQSQVYSFSTDLKLVDTNKVQDY 349

QY 299 -----LPROLMRLAERAFKSNHKTIBPRGNSLKVIGEAS-----FQDNDL 339
Db 350 FKFLQEPFR-LTHVSOQAIPVSATNAVENLNVLKKVKHANLNLSIPTQFNDFYFVNL 408
QY 340 SOLMLPDGLE-----KIESBAFTGNPGDDHYNRNVLMVTKSGKNPQSLATE-NTYVNP 391
Db 409 QHLKLEFGLPILTKQLEN-----LLLSIKQSKNLKFLRLNFYTYAQ 453
QY 392 DKSLWQESPEIDYTKWLBEDFTYQNSVTGFNSKGLQKVRNKNLEIPKQHGNGVITIEIG 451
Db 454 ETSRQK-----ILQOATT-----IKLNKNKQOETPE-----TKDETP 487
QY 452 DNAFRNVDFQNTLARKYDLE-----VKLPST-----IRKIGAFAFOSNLKSPFAS--DDLSEI 504
Db 488 SESTSGMKPFDHLSELTELEDFSVNLQATQEIYDSLHLKILIRSTNLKPKFQKSYKTEMEKS 547
QY 505 KEGAF--MNNRIETL-ELKDKLVITGDAAFHINHIAIVLPESV-----QBI--GR 550
Db 548 KMDTIDLKNIYETLNNLKRCVSNISNP--HGNISVELTNKDSFTYKFKLTNLQELQAHK 605
QY 551 SAFRONGANNLIPMGSKVITGEMAFPLNRLEHLDLSBOKQLETPVQAFSDNALKEVLL 610
Db 606 YTFKQNE-----FQFNVK-----SAKIESSLESLEDIDSLCKSIASCKNLQNVNI 652
QY 611 PASL-----KTIREAFKQHL-----KQLEVASALSHTAFNALDDNDGDEQFNDKVVVKTTHN 664
Db 653 IASLLYPNNIQPNPKNLLFPKQFQKLNLENVSINCILD-----TQPRD 713
QY 665 SYALADGBHFIVDPKLASTIVDLSEKILKIEGLDYSTLRQTTQ-----TQPRD 713
Db 695 QHILNLSIEFFLEKKNKIKAFI--LKYLYLQYLYDTYKFKTLQQLPELNQVYINQOLEE 752
QY 714 MTTAKALLSKSNLQSGEKQK-----FLOBAQFFLRVLDLKAIA--KAELKALVTK 762
Db 753 LTVSE-----VHKQVWENHKQAFYBPLCEFIKESQTLQIDFQNTVSDSDSIKKILESI 808
QY 763 KATKNGQLLERSINKAVLAYNNSATKKANVKELEKELDLTGLVSGKPLAQATWVQGVY 822
Db 809 SESKHYHLRLPSPQ-----SSSLIKSEN-----BEIQELLKACDEKGLVKA-----851
QY 823 LKTLPLPPEYVIGLNVYFD 842
Db 852 YKFPPLCLP-----TGTYD 866

RESULT 52

US-08-974-549A-221
Sequence 221, Application US/08974549A
Patent No. 6166178

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Katen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A

FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-974-549A-221

Query Match 3.3%; Score 167; DB 3; Length 872;
Best Local Similarity 20.5%; Pred. No. 0.0036;
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;

QY 13 TTVSVVTHNQEVF-----SLVKEPIK-----QTQASSISIGADYAESGSKLK 57
Db 17 TNLDFVLQNLVYKQSIHYKTKQQQIKEDLKLKPKNQDQDNGSNDDEENNSN-K 75
QY 58 INETSGPVDDTVTLFSDKRTTPEKID-----NLAKGPBEQLKAVTENTSEKQIT 110
Db 76 QOELLRRVNIQKQVQLIKKYGSKVEKDLNLDENKNGLSRQOVKEEQRLTITEQVK 135
QY 111 SGSOLEQSKESLSLNTKVPSTSNWEICDFITKNGTLVLSKSGVEKLSQTDHLVLPQAA 170
Db 136 YQNLVFNMDYQDLNLSGCHRRHRTDYDT-----EKWFEISH-----174
QY 171 DGTQLIQVAFPTDKKTAIAEYTSRAGENGEISQLDVGKEIINEGE--VFNSYLLKK 228
Db 175 --DQKNVYSIYA---NQKTSYCWMLKDYFNKNYDHLNVSNRLTEAEFYAFDDFSQTI 229
QY 229 VTITGYKHIGQDAFVKNK-----NTAEV-----NLPESET 260
Db 230 KLTNNSYQTVNIDVNFNDNLCLALLRPLLSLRFNLIRSSYTRNOYNPEKIGELLE 289
QY 261 ISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLS-----298

```
Db 290 IFAVFVSHRHLOGIHLQVCEAFQOYLNVSSSQISVKDSQLOVVSFSTDLKLVDTNKVDY 349
Qy 299 -----LPRQLMLAERAFKSNHIKTIIEFRGNSLKVIGEAS-----FQDNDL 339
Db 350 FKFLQEFPR-LTHVSOQAIPVSATNAVENLNLVLLKKVKHANLNLVSIPTQPNFDFYFVNL 408
Qy 340 SOLMLPDGLE-----KIESEAFGNPDGDDHNNRVLVLTSGKRNPSGLATE-NTYVNP 391
Db 409 QHLKLEFGLFNILTKQLEN-----LLLSIKQSKNLKFLRLNFYFYVAQ 453
Qy 392 DKSILWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNLLEIPKQNGVITIEIG 451
Db 454 ETSRKO-----ILKOATT-----IKNLKNNKQOETPE-----TKDETP 487
Qy 452 DNAFRNVDFQNTLRKYDLE-----VKLPST-----IRKICAFQFOSNNLKSPEAS--DDLERI 504
Db 488 SESTSCMKFFDHLSELTELEDFSVNLQATQEIYDSLHLKLLIRSTNLKKFLSVKYEMEKS 547
Qy 505 KEGAP--MNNRIETL-ELKDKLVITIGDAAFHINHIYAVLPSV-----QEI--GR 550
Db 548 KMDTFIDLKNIYETLNNLRKCSVNI SNP--HGNISYELTNKDSFTYFKKLTNLQELQHA 605
Qy 551 SAPRONGANNLFWGSKVTLGEMAFLSNRLEHLDLSEQOLTEIPVQAFSDNALKEVL 610
Db 606 YTFKQNE-----QFNNVK-----SAKIESSLESLEDIDSLCKSIASCNKLQVNI 652
Qy 611 PASL-----KTIREAFKKNHL-----KOLEVASALSHIAFNALDNDGDGEQFDNKKVVKTHN 664
Db 653 IASLLYPNNIKNPFPKNLLEFPQEFQKLNLENVINCILD-----694
Qy 665 SYALADGEHPIVDPKLSTIVOLEKILKILBGLDYSTLRQTTQ-----TQFRD 713
Db 695 QHILNSISFLEKNKKIKAPI--LKRYLLQVYLDYTLFKTLQQLPELNQVYINQOLBE 752
Qy 714 MTAGKALLSKNLROGEKOK-----PLQEAQFFLGRVLDLKAIA---KAERKALVTK 762
Db 753 LTVSE-----VHKQVWENHKKQAFYEPCEFIKESQTLQIDFDQNTVSDSIIKILESI 808
Qy 763 KATQKQLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLVEKGPLAQATWQGVY 822
Db 809 SESKYHYLRNLP SQ-----SSSLIKSEN-----EIQELLKACDEKGVLVKA-----851
Qy 823 LKUTPLPLPEYIGLVNVPD 842
Db 852 YVKFPLCLP-----TGTYD 866

RESULT 53
US-08-854-050-8
; Sequence 8, Application US/08854050
; Patent No. 6261836
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6261836el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-0029930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-854-050-8

Query Match 3.3%; Score 167; DB 3; Length 872;
Best Local Similarity 20.5%; Pred. No. 0.0036;
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;

Qy 13 TTVSVVTHNQEVF-----SLVKEPIK-----QTOASSISGADYAESKSKUK 57
Db 17 TNLDFVLQNLVYKSIQIEHYKTTQOOQIKEEDLKLKFKNQDQDNGSNDDEENSN-K 75
Qy 58 INETSGPVDDTVDLFSDKRTTPEIKD-----NLAQPREQELKAVTENTESEKOIT 110
Db 76 QOELLRRVNLQKQVQLIKKVGSKVEKDLNLEDENKKNGLSEQQVKEQELRTITEQVK 135
Qy 111 SCSQLEQSKESLSLNTKVPSTSNWEICDFITKGNLTVLGSLKSGVEKLSQTDHLVLPQAA 170
Db 136 YQNLVFNMDYQLDLNESGGHRRHRETDYD-----EKWFEISH-----174
Qy 171 DGTQLIQVASPAFTPKDTAIAEYTSRAGENGEISQLDVGKBIINEGE--VFNSYLLKK 228
Db 175 --DQKNYVSIYA--NOKTSYCWMLKDYFNKNYDLNVSINRLETEAEFYAFDDPSQTI 229
Qy 229 VTIPGYKHIGODAFVDNK-----NIAEV-----NLPESET 260
Db 230 KLTNNSYQTVNIDVNFNNLCLALLRFLLSLERNILNIRSSYTRNQVNFKEIGELLE 289
Qy 261 ISDYAFALHAKQIDLPNLKALGELAFDQITGKLS-----298
Db 290 IFAVVFSHRHLOGIHLQVCEAFQOYLNVSSSQISVKDSQLOVVSFSTDLKLVDTNKVDY 349
Qy 299 -----LPRQLMLAERAFKSNHIKTIIEFRGNSLKVIGEAS-----FQDNDL 339
Db 350 FKFLQEFPR-LTHVSOQAIPVSATNAVENLNLVLLKKVKHANLNLVSIPTQPNFDFYFVNL 408
Qy 340 SOLMLPDGLE-----KIESEAFGNPDGDDHNNRVLVLTSGKRNPSGLATE-NTYVNP 391
Db 409 QHLKLEFGLFNILTKQLEN-----LLLSIKQSKNLKFLRLNFYFYVAQ 453
Qy 392 DKSILWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNLLEIPKQNGVITIEIG 451
```


NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/721,456
FILING DATE: 22-NOV. 6617110-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 221:
US-09-721-456-221

Query Match 3.3%; Score 167; DB 4; Length 872;
Best Local Similarity 20.5%; Pred. No. 0.0036;
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;
QY 13 TTVSVVTHNQEVF-----SIVKEPIK-----QTQASSSTSGADYAESSCKLK 57
DB 17 TNLDFVLQNLVYKSGIEHYKTOQQOIKEDLKLKPKNQDQNGSNDDEENNKN-K 75
QY 58 INETSGPVDDTDLFSKRTTPEKID-----NLAKGPREQBLKAVTENSEKQIT 110
DB 76 QQELRRVNIQKQQLIKKVGSKVDLNLNEDENKKNGLSEQQVKEEQLRTITEQVK 135
QY 111 SSGSQLEQSKESLSLNTKPTSTNWIICDPTKGNLTVGLSKSGVKKLSQTHLVLPQAA 170
DB 136 YQNLVFNMDYQDLNLESNGHRRHRETDYD-----EKWFEISH----- 174
QY 171 DGTQLIQVASFAPTPDKKTAIAEYTSRAGENGESQLDQVDGKEIINEGE--VFNSYLLKK 228

DB 175 --DQKNVSYIA---NQTSYCWMLKDYFNKNYDHLNVSINRLETAEFYAFDDPSQTI 229
QY 229 VTIPGYKHGQDAFVNDK-----NIAEV-----NLPESLET 260
DB 230 KLTNNSYQTVNIDVFNNDNLCLALLRFLSLERFNILNRSYTRNQYNPEKIGELLET 289
QY 261 ISDYAFAPHALAKQIDLPDLNKAIGELAFPDNQITGKLS----- 298
DB 290 IPAVVFSHRHLOHQLQVPCFAQYLVNSSQISVKDSQSQVYSFSTDLKLVDNKNQVDY 349
QY 299 -----LPRQLMLAERAPKSNHIKTIEFRGNSLKVIGEAS-----FQNDNL 339
DB 350 PKFLQEPFR-LTHVSQQAIPVSATNAVENLNLVKVGHANLNLVSIPTQNFDFYFVNL 408
QY 340 SQLMLPDGLE-----KIESEAFGNPDGDHNNRVVLMTKSGKNPSGLATE-NTYVNP 391
DB 409 QHLKLEFGLNEPILTKQKLEN-----LLLSIKQSKNLKFLRLNFTYVAQ 453
QY 392 DKSLWQSPEDITYKNLEEDFTYQKNSVTGNSKGLQVKKNKLEIPKQHGVTITIG 451
DB 454 ETSRKQ-----ILKQATT-----IKNLKNNKQEEETPE---TKDETP 487
QY 452 DNAFRNVDFONKTLRKVDLEB--VKLPST---IRKIGAFAPQSNLKSFEAS--DDLREI 504
DB 488 SESTSGMKFPDHLSELTELEDFSVNLQATQEIYDSLHLKLLIRSTNLKPKLSYKEMEKS 547
QY 505 KEGAP--MNNRIETL-ELKOKLVTIGDAAPHINHIYAVLPESV-----QEI--GR 550
DB 548 KMDTFIDLNKNIYETLNNLKRCVSNISNP--HGNTSYELTNKDSIFYKPKLTLNQLQHAQ 605
QY 551 SAFRQNGANNLIFMGSKVKTGEMAFPLSNRLEHLDSEQKQLTPIVQAFPSDNALKEVLL 610
DB 606 YTFKQNE-----QFNNVYK-----SAKIESSESLESDIDSLCKSIASCKNLQNVNI 652
QY 611 PASL---KTIREAPKKNHL---KQLEVASALSALHAFNALDNDGDEQFDKNVVKTHN 664
DB 653 IASLLYPNNIQKNPKNPKNLLFPKQFQOLKNLENVINCILD----- 694
QY 665 SYALADGEHFIVDPDKLSSTIVLEKILKLEGLDYSTLRQTQ-----TQPRD 713
DB 695 QHILNSISEFLEKKNKIKAFI--LKRYVLLQYLYDYTKLFTLQQLPELNQVYINQOLEE 752
QY 714 MTTAGKALLSKSNLRQGEKQK-----FLOEQAQFPLGRVLDLKAIA--KAERKALVTK 762
DB 753 LTVSE---VHKQWENHKQKAFYEPLCEFIKESSTQLQLIDFQNTVSDSDSIKKILESI 808
QY 763 KATKNGQLLERSINKAVLAYNNSAIKCANVRLKELDLTLGLVEGKGLAQATWQGVY 822
DB 809 SESKYHYLRLNPSQ-----SSSLIKSEN-----EIQELLKACDEKGLVYKA----- 851
QY 823 LLKTPPLPEYYIGLVNLYVD 842
DB 852 YKFPPLCLP-----TGTYYD 866
RESULT 59
US-09-766-253-8
; Sequence 8, Application US/09766253
; Patent No. 6808880
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin
; Andrews, William H.
; TITLE OF INVENTION: No. 6808880el Telomerase
; NUMBER OF SEQUENCES: 171
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/766,253
 FILING DATE: 19-Jan-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/846,017
 FILING DATE: 1997-04-25
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002920US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 US-09-766-253-8

Query Match 3.3%; Score 167; DB 4; Length 872;
 Best Local Similarity 20.5%; Pred. No. 0.0036;
 Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;

QY	13	TTVSVVTHNOEVP-----SLVKEPIK-----QTQASSISGADYAESSGSKLK 57
DB	17	TNDFVLQNLLEVYQIEHYKTKQQQIKBEDLKLKFKNQDQDGNDDDEENNNS-K 75
QY	58	INETSFPVDVTDFLPSDKRTTPEKID-----NLAKGPRQELKAVTENTSEKQIT 110
DB	76	QOELLRRVNIKQOVQLIKVGSKEVDNLNEDENKNGLSQOVKEQLRTITEQVK 135
QY	111	SGSQLEQSKESLSNKTVPSTSNWEICDFITKNTLVGLSKGVEKLSQTDHLVPSQAA 170
DB	136	YQMLVFNMDYQLDLNESGGHRRHRETVDYD-----EKWFEISH----- 174
QY	171	DGTQLIQVASFATPDKTKTAIAEYTSRAGENGESOLDVDGKEIINEGE--VPNSVLLKK 228
DB	175	--DQKNVSTIYA---NQTSYCWLKDYFNKNYDHLNVSINRLTEAEFYAFDDFSQTI 229
QY	229	VTIPTGYKHIGQDAFVDNK-----NIAEV-----NLPSLET 260
DB	230	KLTNNSYQTVNIDVFNDDNLCLALLRFLLSLRFNLIARSSYTRNQYNFEKIGELLET 289
QY	261	ISDYAFALHALKQIDLPDNLKATGELAFPDNQITKGLS-----FQNDL 339
DB	290	IFAVFESHRLQGIHLQVPCFAQYLVNSSQSISVKDSQOVVVSFSTDLKLVDTNKVQDY 349
QY	299	-----LPRQLMELAEAFKSNHIKTIIEPRGNSLKVIGAS-----FQNDL 339
DB	350	FKFLQBFPR-LTHYSQQAIPVSATNAVENLVLLKKVKHANLVLVSIPQFNDFDFVNL 408
QY	340	SQMLMPDGLG-----KIESEFTGNGDDHYNNRVWLTKSGKNPSGLATE-NTYVNP 391
DB	409	QHLKLEFGLPEPNTLTQKLEN-----LLLSIKOSKMLKFLRLNFTYVVAQ 453
QY	392	DKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPKHQNGVTITETG 451

DB	454	ETSRKQ-----ILKQATT-----IKNLKNNKNOBETPE-----TKDETP 487
QY	452	DNAFRNVDFQNTLRKYDLEE--VKLPST---IRKIGAFAFOSNNLKSPFAS--DDLEI 504
DB	488	SESTSGMKFPDHLSELTELEDFSVNLQATQEIYDSLHKLLIRSTNLKPKLSYKYEWEKS 547
QY	505	KEGAF--MNNRIETL-ELKDKLVITIGDAAFHINHIYAIYVLPESV-----QEI--GR 550
DB	548	KMDTFIDLKNIYETLNNLKRCVSNISNP--HGNISYELTNKDSFTFYKPKLTNLQELQHA 605
QY	551	SAFRONGANLIPMGSKVKTGEMAFLSNRLEHLDLSEQKQLTETPQVAFSONALKEVLL 610
DB	606	YTFKQNE-----FQFNNVK-----SAKIESSSLESLEDLSLCKSIASCNKLNQVNI 652
QY	611	PASL---KTIREAFKQNH---KOLEVASALSHIAFNALDNDGDGEQDFDNKVVVKTHN 664
DB	653	IASLLYPNNIQKNPKNPKNLLFFKQFEQKNUENVSINCILD----- 694
QY	665	SYALADGEHFIVDPDKLSSTIVDLKILKLIBGLDYSTLROTTQ-----TQFRD 713
DB	695	QHILANSISSEFLEKKNKIKAFI--LKRYLLQYVLDYTKLFTLQQLPELNQVYINQLEE 752
QY	714	MTTACKALLSKSNLRQGEKQK-----FLOBAOFFLGRVLDLKAIA---KAEKALVTK 762
DB	753	LTVSE-----VHKQVWENHKQAFYEPCEFIKESQTLQIDFDQNTVSDSDSIKKILES 808
QY	763	KATKNGQLLERSINKAVLAYNNSAIKKANVKELEKELDLTGLVSGKGPLAQATWVQGY 822
DB	809	SESKYHHYLRNPSQ-----SSLSIKSEN-----BEIQELLKACDEKGVLVKA----- 851
QY	823	LLKTPPLPPEYVIGLVNVPD 842
DB	852	YVKEPLCLP-----TGTYYD 866

RESULT 60
 US-09-766-253-54
 ; Sequence 54, Application US/09766253
 ; Patent No. 6808880
 ; GENERAL INFORMATION:
 ; APPLICANT: Cech, Thomas R.
 ; Lingner, Joachim
 ; Nakamura, Toru
 ; Chapman, Karen B.
 ; Morin, Gregg B.
 ; Harley, Calvin
 ; Andrews, William H.
 ; TITLE OF INVENTION: No. 6808880el Telomerase
 ; NUMBER OF SEQUENCES: 171
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 94111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/766,253
 ; FILING DATE: 19-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/846,017
 ; FILING DATE: 1997-04-25
 ; APPLICATION NUMBER: US 08/724,643
 ; FILING DATE: 01-OCT-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Apple, Randolph T.
 ; REGISTRATION NUMBER: 36,429

Db 640 ILALEKEKQLOKQLOALTSRKAILKKAQEKERHLREELKQKODDYNRLQEQFQDSKE 699
QY 433 NKMLEIPKOHNGVTITIEIGNAPRNVDFOKTLRKYDLDEEVLKPLSTIRKIGAPAFOSN-- 490
Db 700 NENIGDQLQLOIQVRESIDKLPSTDDQESCSSTGLBEPLFKATEQHTQPVLESNLC 759
QY 491 -----NLKSPFASDDLEIEKEGA-----FMNRIETLELKD 522
Db 760 POMPSSHESDASALQCGTSVAQIAKQALKEIEAEKVELELVSTSTELTKKSEVFLQEQ 819
QY 523 LVITIG---DAAPHINHIYAIVLPESVQE-IGRAFRQNGANNLIFMGSKVKTL----- 571
Db 820 INKQGLEISLKTQVSH-EAEVHAESLQQLLESQLOAGLEHRELPKLDQLKLSKX 878
QY 572 -GEMAFLSNRLE-----HLDLSEKQOL-----TEIPVQAFS-DNALKEV----- 608
Db 879 EEDVSLSQSLSEKAALTKIOTIEIEQEDLIKALHTQLEMAQEKEDERIKQLOVELCEM 938
QY 609 -----LLPASLKTIRREAPKQNHKLQLEVASALSHI--AFNALDDN 647
Db 939 KOKPEBIGESRAKQIQKQLQAALIS-RKEALKENKSLQEELSLARGTIERLTKSLADV 997
QY 648 DGDGEQDNK-----VVVTHNSYALADGEHFIVDPDKLSST-----IVDL 688
Db 998 ESQVSAQNEKQTVLGRLLALQBERDKLITEMDRSLLENQSLSSCESLKLALGLTEDK 1057
QY 689 EKILKLEGLDYSLRTQT--OTQPRDMTTAGKALLSKNLRQEKQK-----LOEA 739
Db 1058 EKLKVEIESLSKIAESTEWQEKHELOKEYBILLOSTVENVSNEAERIQHVAVRQEK 1117
QY 740 QPFLGRVLDLDAIAK-AEKALVTKATKQQLLERSINKAVLAYNNNSAIKKANVRKELE 798
Db 1118 QELYGLKLRSTANKETEKQOL--QEAQEQEMEEMKEMKRFKAS-----KQOKILELEE 1169
QY 799 LDLLTLGLVGKGLAQATVWQGVYLLKPLPLPEYIYGLNVFDPKSGKLIYALDMSDTIG 858
Db 1170 NDRLEAHPAGDTAKECHET-----LLSSNASMKEBELERKMEYETLSKRFQSL-MS- 1221
QY 859 EGQKDAYGNPILNVDDNDEGYHALAVATLADYEGLDIKTLNLSKLSQTSIRQVPTAAVH 918
Db 1222 --BKDSLSEVQDLKHQIEG-----NVSKQANLEATE-----KHNDQNTVEEGT- 1264
QY 919 RADIFQAIQAAAEABQL-LPKPGTHSEKSSSESAN---SKD 957
Db 1265 ----QSIPEETEEOQSLSNSTRTCTSESVPSSAKSANPAVSKD 1302

RESULT 62
US-09-583-110-4036
; Sequence 4036, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4036
; LENGTH: 1078
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4036

Query Match 3.3%; Score 166.5; DB 4; Length 1078;
Best Local Similarity 19.0%; Pred. No. 0.0053;

Matches 226; Conservative 172; Mismatches 431; Indels 359; Gaps 59;
QY 4 HLKTVALTUTTVSVVTHNQEVSFLVKEPIKLTQOASSISGADYAE-----SSGSKSKLIN 59
Db 35 HWMTVGAETTTNTTQQSQ-----KEVOYQORDTKNLVENGDFGQTEGSSPWTGSKAQ 88
QY 60 ETSQVPDVTVDLFSQKRTTPPEKIKDNLAKGRPEQELKAVTENTSEKOITSGSOLEQSK 119
Db 89 GWSWVDQKNS--ADASTRVTEAKDG-----AIT-----ISSPEKLR--- 124
QY 120 ELSLNTKTVPSSTNWEI--CDFITKGNLTVGLSKSGVSKLSQTDHLVLPQAADGTOLIQV 178
Db 125 --AAVHRMVPPIEAKKKYKLPFKITKDNKVGIARVRIIEESGDKRLMNSATTSCTKDWQT 182
QY 179 --ASFAPTDKKTAATAEYTSRAGEISOLVDGKEI---INEGVFNSYLLKKVTIPT 233
Db 183 IEADYSPITLDVVKIKLELFEYTG-TGTVSFKDIELVEVADQSEDSQTDKQLEEKIDLP 241
QY 234 GYKHIGQDAFVDNKNIAEVLNPESETISDYAFALHALKQIDLPDLNKAIGELAFPDNQI 293
Db 242 GKXHV-----FSLADYTY-----KVENPD-----VASVKNGI 268
QY 294 TGKL-----SLPQLMLRAERAPKSNHIKIEFRGNSLKVIGEASFQD 336
Db 269 LEPLKEGTTNVIVSKDGEVKKIPKILASVKDAYTA---RLDDWNG---IAGNQYVD 321
QY 337 NDLSQL-----MLPDGLEKIESEFTGNPGDDHNNRVVLTWTKSGKNPSGLATENT 387
Db 322 SKNEQMAKLNQBLEKQVADSLSSISQA-----DRTYLMKFSNYKMSANLTAT 370
QY 388 Y-----VNPQKSLMQESPEI-----DYTKWLEEDFTYQKNSVTG----- 421
Db 371 YRKLEMAKQVTPSSRYYYQDETIVRTVDSMEWHKHVYNSEKSI VGNWMDYEIGTPRA 430
QY 422 -----PSNKGLOKVKR--NKNLEIPKQHGNTVITETGDNAFR-----NV 458
Db 431 INNTLSLMEKYEFSDEBEIKKYTVIEKFVPDPPEFRKTT-----DNPFKALGGNLVDMGRV 485
QY 459 DFQNKTLRKDYLEEVKLPSTIRKI-----GAFAFOSNNLKSFEASDDLEIEKEGAF 509
Db 486 KVIAGLLRKDDQ-----ISSTIRSIQVFKLVQGGEGFYQDGSY-----IDHTNVATGAY 537
QY 510 MNRIETLEL-----KOKLVITIGDAAFH--INHIYAIVLPESV---QEIGRSA 552
Db 538 GNVLDLGLSQLPLVIQKTKNPIDKQKQTM---YHWIDKSFAPLLVNGELMDMSGRSI 593
QY 553 FRQNGANNLIFMGSKVKTLGEMAFLSNRLEHLDLSEKQKLTETIPVQAF--SDNALKEVLLP 611
Db 594 SRANSEGHV---AAVEVLRTGI---HRIADMSEGETKQRLQSLVKTVIQSDSYVDVF-- 643
QY 612 ASLKTIRREAFKKNHLKQLEVASA--LSHI--AFNALDDND-----GDEQFDNKV 658
Db 644 KNLKTYKDISLMQSLSDAGVASVPTSYLSAFNMKMDKTAMYNAEKGFGLSLFSSRTL 703
QY 659 VKTHN-----SYALADGEHFT-----VDPDKLSSTI-----VDLEK 690
Db 704 NYEHNNKENKRCWYTSDCGMFYLYNGDLSHYSDGYNFTVNPYKMPGTETEDAKRADSDTK 763
QY 691 ILKLIEGLDYSLRTOTQTFQPRDMTTAGKALLSKNLRQEKQK--LQBAQFFLGRVDLD 749
Db 764 VLP-SAFVGTSKLDDANATATMDFTNWNQTLTA-----HKSFWMLKDKIAFLG----- 810
QY 750 KAIAKAEKALVTYKATKNGCOLLERSINKAVLAYNNNSAIKKANVRKLEKELDLTLGLVECK 809
Db 811 ---SNIQNTSTDTAATTIDQRKLESSNPYKVVND---KEASLTBEQKDY----- 854
QY 810 GPLAQATWQGVYLLKTLPLPEYIYGLNVYFDKSGKLIYALDMSDT-----IGEQ 861
Db 855 -PETQSVFLESSDSKKN-----IGY-FFPKKS-----SISMSKALQKAWKINEQ 899
QY 862 KD-AYGNPILNVDE-----DNEGYHALAVATLADYEGLDIKTLNLSKLSQTSIRQVPT 914
Db 900 SKKEVNEFLTISQAHKQNGDSYGYMLIPNVDRATFNQW--IKELESSLIENNETLQSVVD 958

QY 915 AAYHRAGI-----FOAIONAA-----ABEQLLPKPGTHS- 944
DB 959 AKQGVWGVKYVDDSVSTISNQFVLKRGVYTIKREGEYKIAYNPTEQESAPDQEVFK 1018
QY 945 -EKSSSSSANSKDRGLQSNKPTNRRHSAILPRTGSGSFVYGLGY 991
DB 1019 LEQAAQPOVQNSKEKEKEEKNHSDQKN--LPQTG-BQQSILASLGF 1063

RESULT 63
US-09-107-433-4843
; Sequence 4843 Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4843:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1080 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1080
; SEQUENCE DESCRIPTION: SEQ ID NO: 4843:
US-09-107-433-4843

Query Match 3.3%; Score 166.5; DB 4; Length 1080;
Best Local Similarity 19.0%; Pred. No. 0.0053;
Matches 226; Conservative 172; Mismatches 431; Indels 359; Gaps 59;

QY 4 HLKVALTLTVTVVTHNQEVSLVKEPIIKQTQASSISGADYAE-----SSGKSKLXIN 59
DB 37 HTWVGAEETTTNTIQSQ-----KEVQYQQRDTKNLVENGDFQGTEDGSSPWTGSKAQ 90
QY 60 ETSQGVDDTVTLDFSDKRTTPBKIKDLAKGPREQELKAVTENTSEKQITSGSQLEOSK 119
DB 91 GWSTWVDQKNSS--ADASTRVIEAKDG-----AIT-----ISSPEKLK--- 126

QY 120 ESLINKTVPTSTNWEI--CDPITKNTLVGLSKGVKLSQTDHLVLPQADGTQLIQV 178
DB 127 --AAVHRVPIEAKKVKLRKFKIKTDNKKVGIKRVIIIEESGKDKRLMNSATTSQTKDWOT 184
QY 179 --ASFAFPTDKKTAIAEYTSRAGENGEISQLDVQKEI---INEGEVFNLSLLKKVTPPT 233
DB 185 IEADYSPTLDVDKIKLELFPYETG--TGIVSPKDIELVEVADQLSEDSQTKQLEBKIDLP 243
QY 234 GYKHIGQDAFVDNKNIABVNLPESETISDYAPFAHLALKQIDLPDLNKAIGELAFPDNQI 293
DB 244 GKCHV-----FSLADITY-----KVENPD-----VASVKNGI 270
QY 294 TGKL-----SLPQLMRLAERAPKSNHIKTIIEFRGNSLKVIGEASFQD 336
DB 271 LEPLKRGTTNVIVSKOGKEVKKIPKLILASVKDAYTA---RLDDWNG---IIAGNQYD 323
QY 337 NDLSQL-----MLPDGLEKIESAFTGNPDGDHNNRVVLWTKSGKNPSGLATENT 387
DB 324 SKNEQMAKLNQELSGKVADSLSSISSQA-----DRTYLWEFSNYKMSANULTAT 372
QY 388 Y-----VNPDKSLMQSPBI---DYTKWLBEDFTYQKNSVTG----- 421
DB 373 YRKLEEMAKQVTNPSSRYQDETIVRTVRDSMEWMHKHVYNSEKSI VGNMWDYEIGTPRA 432
QY 422 -----FSNKGLQKVKR--NKNLIEIPKHNGVTIIEIGDNAPR-----NV 458
DB 433 INNTLSLMKEYFSDEEIKKYTDVIEKFVDPPEHFRKTT-----DNPFKALGNLVDMGKV 487
QY 459 DFQNTLRKYDLEEVKLPSTIRKI-----GAFAPQSNLKSFEASDLEBEKEGAF 509
DB 488 KVIAGLLKQDOE---ISSTIRSEIQVFKLVQOEGEFGYQGSY-----IDHTNVAYTQAY 539
QY 510 MNNRIETLEL-----KDKLVITIGDAAPH-INHIYAIVLPESV---OEIGBSA 552
DB 540 GNVLIDGLSQLPVIQTKNPKIDKQKQTM---YHWIDKSFAPLLVNGELMDMSGRSI 595
QY 553 FRQGANNLIFMGSKVKYTLGEMAFLSNLEHLDSSEQOLTEIPIQAF--SNALKEVLLP 611
DB 596 SRANSEGHV-----AAVEVLRI-----HRIADMSGETKQRLQSLVKTIVQSDSYVDV 645
QY 612 ASLKTIREAFKKNHLKQLEVASA--LSHI-AFNALDDND-----GDEQPDNKKV 658
DB 646 KNLTKYKDISLMQSLSDAGVASVPTSYLSAFKNKMDKTAMNAEKGFGLSLFSSRTL 705
QY 659 VKTHN-----SYALADGEHPI-----VDPDKLSSTI-----VDLEK 690
DB 706 NYEHMKNKRGWYTSQGMFYLYNGDLSHYSDGYWPTVNPYKMPGTTTDAKRASTDGK 765
QY 691 ILKLEGLDYSTLRQTTQTFQDMTTACKALLSKNLRQGEKQKF--LOEAQFFLGRVLD 749
DB 766 VLP--SAFVGTSKLDDANATATMDFTNMNQTUTA-----HKSWFMLKDKIAFLG----- 812
QY 750 KAIKAEKALVTKATKNGQLLERSINKAVLAYNNSAIKKAANVKRLEKELDLTLGLVEGK 809
DB 813 ---SNIQNTSTDTAATTIDQRLKLESSNPYKVYND---KEASLTEQEKDY----- 856
QY 810 GPLAQATWQGVYLLKTPLEPLPEYIYGLNVYFDKSGKLIYALDMSDT-----IGBGQ 861
DB 857 -PETQSVFLESSDSKCN-----IGY-FFFKKS-----SISMSKALQKGAWKDINEGQ 901
QY 862 KD-AYGNPILNVD-----DNEGYHALAVATLADYEGLDIKTILNSKLSQTSIRQVPT 914
DB 902 SDKEVENSEFLTISQAHKQNGDSYGYMLIPNVDRATFNQM--IKELESSLIENNETLQSYVD 960
QY 915 AAYHRAGI-----FOAIONAA-----ABEQLLPKPGTHS- 944
DB 961 AKQGVWGVKYVDDSVSTISNQFVLKRGVYTIKREGEYKIAYNPTEQESAPDQEVFK 1020
QY 945 -EKSSSSSANSKDRGLQSNKPTNRRHSAILPRTGSGSFVYGLGY 991
DB 1021 LEQAAQPOVQNSKEKEKEEKNHSDQKN--LPQTG-BQQSILASLGF 1065


```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-496A-10

Query Match 3.3%; Score 166; DB 2; Length 2366;
Best Local Similarity 20.7%; Pred. No. 0.019;
Matches 186; Conservative 129; Mismatches 304; Indels 278; Gaps 49;

QY 20 HNEVFSLVKPKILKQTSISGADYABSGSKLKNINSGVDPVTVTLDFS---DK 76
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
330 YTSFHDMLDEEV---QSSPESVL-----ASKSDKS-----EIFSSLGDM 366
QY 77 RTTPEKIKDNL-AKGPQELKAVTENTSE---KQITSGSQLEQSKESLSINKTVPTS 132
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
367 EASPLEVKIAFNSKGIINQGLISVXDSCSNLIVKQI-----ENRYKILNLSNLP 419
QY 133 NWEICDFITKGNLTGLSKSGVEKLSQTDHLVLPQAAADGTOLIOVASF---AFTPKKT 189
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 --EDNDFNTTNTFI-----DSINAEANADNGRPMELGKYLKRVGFFPDVKT 464
QY 190 AIAEYTSRAGENGESQLDVGKE-----IINEGEVNSYLLKKVITPTGYKHG 239
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
465 TI-----NLSQPEAYAAAYQDLLMFKEGSM-NIHLI-----494
QY 240 QDAFVKNKNIARVNLPESLETISDYAPAHALAKQIDLPNKLKAIGELAFFDQNTGKLSL 299
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
495 -EADLRFNEIKTNSISQSTE-----QEMASLWSFDD-----524
QY 300 PQLMLRAERAPKSNHKTIEPRGNSLKVIGASFDQN-DLSQLMLPDG---LEKIESEA 355
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
525 ARAKAQFEE--YKRYN-----FEQS-----LGE---DDNLDFSQNIWVKYLEKISSLA 570
QY 356 FTGNPGDDHNNRVVLTWTKSGKNPGLATENTYVNP--DKSIWQBSPEIDYTKWLEEDFT 413
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
571 RSSERYIHY-----IVQLQGDKISYEACNLFAKTPYDSVLFQKNIE-----DSEIA 618
QY 414 YKNSVTGFSNGLQKVKENKNLEIPKHONGVITTEIGDNAPRVDFQNKTLRKVDLE-- 471
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
619 YYNPGDG-----EIQEDIKYKIPSIISDRPKLTFIGHG-----KDFNNTDIPAGFDVDSL 671
QY 472 --EVKLPSTIRK--ICAFQFQNNL--KSFEASDDLEEKEGAFMNNRIETLELKD--- 522
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
672 STEIEAADLAKEDISPKSIEINLLGCNMFYSINVEETYPGKLL-----LKVKDKISE 725

;
; RESULT 66
; US-08-915-136-10
; Sequence 10, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989

```

ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-10

Query Match 3.3%; Score 166; DB 3; Length 2366;
Best Local Similarity 20.7%; Pred. No. 0.018;
Matches 186; Conservative 129; Mismatches 304; Indels 278; Gaps 49;

QY 20 HNOEVFSLVKEPILKQTOASSISGADYABSSGSKLKNINETSQVDDTVTDLFS---DK 76
Db 330 YTEHFHMDLDEEV---QSSPESVL---ASKSDKS-----EIFSSLGDM 366
QY 77 RTTPEKIKDNL-AKGPREQLKAVTENTSE---KOITSGSQLEQSKESLSLNKTVPS 132
Db 367 EASPLEVKIAFNKGIINOGLISVXSYCSNLIIVKQI-----ENRYKILNNSLNPAIS 419
QY 133 NWEICDFITKGNLTAVGLSKSGVEKLSQTDHLVLPQAADGTQLIQVASF---AFTPKKT 189
Db 420 --BNDNFNTTTFI-----DSIMAEANADNGRFWELGKYLVRVGFDPDKT 464
QY 190 AIAEYTSRAGENGESISOLDVDGKE-----IINEGEVFNFSYLLKKVTIPTGYKHG 239
Db 465 TI-----NLSGPEAAVAYQDLLMFKESGM-NIHLI-----494
QY 240 QDAFVNKNIAEVLNPLESETISDYAFALHAKQIDLPDLNKAIGELAFDNOITGKLSL 299
Db 495 -EADLRNFEISKVNSQSTE-----QEMASLWSFDD-----524
QY 300 PRLMLRAERAPKSNHIKTIFRGNLSKLVIGEASFDON-DLSQMLPDG---LEKIESEA 355
Db 525 ARAKAQFEE---YKRN-----FEGS-----LGE---DONLDFSQNIIVVDKEYLEKISSLA 570
QY 356 FTQNPQDDHYNRNVWLTKSGKPNGLATENTYVNP--DKSLWQESPEIDYTKWLEEDFT 413
Db 571 RSSERGVHY-----IVQLQDKISYEACNLFAKTPYDSVLFXNIE-----DSETA 618
QY 414 YQKNSVTGFSNKGQVKRNKNLEIPKQNGVVTITIGDNAFNRVDFQNKTLKPYDLE-- 471
Db 619 YVYNPGDG---BIQIDKYKIPSIISDRPKIKLTFIGHG---KDEFNTDIFAGFDVDSL 671
QY 472 --HVKLPSITRK--IGAPAFQSNL--KSPFASDDLEEIKEGAPMNRITETLEKDK--- 522
Db 672 STEIEAIDAKEDISPKSIEINLLGCMNFSYINVEETYPGKLL-----LVKDKKISE 725
QY 523 ---LVTIGDAAPHIN-----HIYAVLPES---VQBIGRSAPRQ-NGANN 560
Db 726 LMFPSISQDSIIIVSANQYEVINSEGRRELLDHSGEWINKEESIIKDISKEYISFPKEN 785
QY 561 LIPFGSKVKTLEGMFLSNRLEH-----LDLSEQQLTEIPVOAFSD---NALKEVLLP 611
Db 786 KITV--KSNLPELSTLLQBIARNNSSDIELEKVMKLTCEINVISNIDTQIVEERIE 843
QY 612 ASUKTIREAFKQNLKQLE-VASAL-----SH-IAFNALDDNDGDQEPQNVV 658
Db 844 AKNLTSDSINYIKDEFKLTIESIDALCDLKQKQNELEDHSFISFE--DISETDEGFSIRFI 901
QY 659 -----VKTHNSVALADGEHFI VDPDKLSSTIVLEKILKLI EGLDYSTLRQT 706
Db 902 NKETGESIFVETKTLFSEYA-----NHITIEISKIGTIFD-----938
QY 707 TOTQFRDMTTAGKALLSKNSLRQGEKQKFLQEAQFFLGRVLDLDAKAEKALVTTKATK 766

Db 939 -----TVNGK-LVKKVNLDTTHEVNTLNAAFFIQSLIEYNSKESLSNLSVAMKVQV 989
QY 767 NGQLLERSINKAVLAYNNSAIKKNVRLKLEKLDLLTGLVEGKGPLAQATWVGYYL 823
Db 990 YQALPSTGLNTTDA---AKVVELSTALDETIDLLPTLSEGL-PIL-ATIIDGVSL 1041
RESULT 67
US-08-957-310-10
Sequence 10, Application US/08957310
Patent No. 6365158
GENERAL INFORMATION:
APPLICANT: Williams, James A.
APPLICANT: Kink, John A.
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
TITLE OF INVENTION: DISEASE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-OCT-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-310-10

Query Match 3.3%; Score 166; DB 3; Length 2366;
Best Local Similarity 20.7%; Pred. No. 0.018;
Matches 186; Conservative 129; Mismatches 304; Indels 278; Gaps 49;
QY 20 HNOEVFSLVKEPILKQTOASSISGADYABSSGSKLKNINETSQVDDTVTDLFS---DK 76
Db 330 YTEHFHMDLDEEV---QSSPESVL---ASKSDKS-----EIFSSLGDM 366
QY 77 RTTPEKIKDNL-AKGPREQLKAVTENTSE---KOITSGSQLEQSKESLSLNKTVPS 132
Db 367 EASPLEVKIAFNKGIINOGLISVXSYCSNLIIVKQI-----ENRYKILNNSLNPAIS 419
QY 133 NWEICDFITKGNLTAVGLSKSGVEKLSQTDHLVLPQAADGTQLIQVASF---AFTPKKT 189

QY 523 -LVTIGDAAPHIN-----HIYAVLPES--VOEIGRSAPRQ-NGANN 560
Db 726 LMPISQDSIIVSANQYEVNRINSEGRRELLDHSGEWINKESIIKDSSKEYISFNPKN 785
QY 561 LIFMGSKVKTGEMAFLSNRLEH-----LDLSEKQLTEIPVOAFSD---NALKEVLIP 611
Db 786 KITV--KSKNLPSTLLQEIIRNNSNSDIELEEKVMLTECEINVISNIDTQIVERIE 843
QY 612 ASLKTIREEAFKXNHLKOLE-VASAL-----SH-IAFNALDDNDGDEQFQPNKV 658
Db 844 AKNLTSDSINYIKDEFKLTIESIDALCDLKQONELEDSHFISFE--DISETDEGFSIRFI 901
QY 659 -----VKTHNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLRQT 706
Db 902 NKETGESIFVETEKTIIFSEYA-----NHITEISKIKGTIFD----- 938
QY 707 TOTQFRDMTTAGKALLSKNSRQGEKQKFOBAQFFLGRVDLDKATAKAELVTKATK 766
Db 939 -----TVNCK-LVKKVNLDTTHEVNTLNAAFFIQSLIEYNSKESLSNLSVAMKVQV 989
QY 767 NGOLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLVEGKPLAQATMVGQVYL 823
Db 990 YAQLFSTGLNTITDA---AKVVELVSTALDETLIDLLPTLSEGL-PII-ATIIDGVSL 1041

RESULT 69

US-09-084-517-10
; Sequence 10, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; 'COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPDH-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-084-517-10

Query Match 3.3%; Score 166; DB 4; Length 2366;
Best Local Similarity 20.7%; Pred. No. 0.018;
Matches 186; Conservative 129; Mismatches 304; Indels 278; Gaps 49;

QY 20 HNOEVFSLVKPEILKQTOASSISGADYAESSGSKLKINETSQDVTDTLDFS---DK 76
Db 330 YTEHFMDLDEEV--QSSPESVL-----ASKSDKS-----BIFSSLGDM 366
QY 77 RTTPKIKIDNL-AKSPROELKAVTENTESE---KQITSGSQLEQSKESLSINKVTPSTS 132
Db 367 EASPLEVKIAFNKSGIINQGLISVKDSCYNLIVKQI-----ENRYKILNLSNPAIS 419
QY 133 NWEICDFITKGNLTAVLSKSGVEKLSQTDHLVLPQAAADGTOLIOVASF---AFTPDKKT 189
Db 420 --EDNDFNTTNTFI-----DSIMAEANADNGRFMMELGKYLVRGVFPDVKI 464
QY 190 AIARYTSRAGENGISQLDVDGKE-----IINEGEVFNYSYLLKKVTIPTGYKHIG 239
Db 465 TI-----NLSGPEAYAAAYQDLLMFKESGM-NIHLI----- 494
QY 240 QDAFVDNKNIAEVNLPESLETISDYAFALHAKQIDLPDLNKAIGELAFPDNOITGKLSL 299
Db 495 -EADLRNFEISKTNISQSTE-----QEMASLMSFDD----- 524
QY 300 PQLMELAEAFKSNHKTIEPRGNLSKVIGEASQDN-DLSQMLPDG---LEKTESEA 355
Db 525 ARAKAQFEE--YKRN-----PEGS-----LGE---DDNLDPSQNVVDKEYLEKISSLA 570
QY 356 FTGNPGDDHNNRVVLTSGKNPSGLATENTYVNP--DKSIWQSPESPEIDYTKWLEEDFT 413
Db 571 RSSERGIHY-----IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIE-----DSEIA 618
QY 414 YOKNSVTGFSNGKLGKVRKNRKNLEIPKQHNGVTTIEIGNAPRVNDVFNKTLRKVDLE-- 471
Db 619 YYNPGDG---EIQEDIKYKIPSIISDRPKIKLTFIGHG---KDEFNIDIFAGFVDLSL 671
QY 472 --EVKLPSTIRK--IGAPAFQSNL--KSFASDDLEELKEGAFWNNRIETLELKKD--- 522
Db 672 STEIEAAILAKEDISPXSIEINLLGCNMFYSINVEETYPGKLL-----LKVKDKISE 725
QY 523 -----LVTIGDAAPHIN-----HIYAVLPES--VOEIGRSAPRQ-NGANN 560
Db 726 LMPISQDSIIVSANQYEVNRINSEGRRELLDHSGEWINKESIIKDSSKEYISFNPKN 785
QY 561 LIFMGSKVKTGEMAFLSNRLEH-----LDLSEKQLTEIPVOAFSD---NALKEVLIP 611
Db 786 KITV--KSKNLPSTLLQEIIRNNSNSDIELEEKVMLTECEINVISNIDTQIVERIE 843
QY 612 ASLKTIREEAFKXNHLKOLE-VASAL-----SH-IAFNALDDNDGDEQFQPNKV 658
Db 844 AKNLTSDSINYIKDEFKLTIESIDALCDLKQONELEDSHFISFE--DISETDEGFSIRFI 901
QY 659 -----VKTHNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLRQT 706
Db 902 NKETGESIFVETEKTIIFSEYA-----NHITEISKIKGTIFD----- 938
QY 707 TOTQFRDMTTAGKALLSKNSRQGEKQKFOBAQFFLGRVDLDKATAKAELVTKATK 766
Db 939 -----TVNCK-LVKKVNLDTTHEVNTLNAAFFIQSLIEYNSKESLSNLSVAMKVQV 989
QY 767 NGOLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLVEGKPLAQATMVGQVYL 823
Db 990 YAQLFSTGLNTITDA---AKVVELVSTALDETLIDLLPTLSEGL-PII-ATIIDGVSL 1041


```
Db 677 MENEKLVACEDVRHOLEECLAGNQLSL-EKNTIV---ETLMEKEIEAEELCWAKRL 732
QY LEEDFTYQK-----NSVTGFSNKGLOKVRKNLEIPKQHGVTITEIGDNAP 455
Db 733 LEEANKYKTIIELSNARNLNTSALQLEHEHLIKINQKDMELAEKKNI---EQMDTQH 789
QY 456 RVNDFONKTLRKVDLEBK-LPSTIRKIGAFAPQSNLKSFEASDDLEIRKEGAFMNNRI 514
Db 790 K-----ETKDLVSSLEQKLTQINKKEIF-TEKLKERSSKLOEBELDKYSQ-ALRKNBI 843
QY 515 --ETLEBKDKLVITIGDAAHINHIYAILVPESVOEIGRSAPFRQNGANNLIFMGSKVKTIG 572
Db 844 LKQITIEKOR--SLGSMKBNNH-----LQEELELRERESQRTAPVADPKTLDVSTELAS 896
QY 573 EMAPLSNRLHLD-----LSEOK-----QLTEIPVQ 598
Db 897 EYSQNTIKHELEEBEIKHKQIIEDQNSQWQLLSQLOEQKEMDFRYQHEQMATHYQ 956
QY 599 AF--SNALKEVLLPASLTKTIREAPKKNHLKQLEVASALSHTAFNALDDNDGDEQDNK 656
Db 957 LFLEKDEEIKS--LQKTIEQIKTLHEERQDIQTDNSDIFQETKVQSLNIENGSEKHD-- 1012
QY 657 VVVKTHNSVALADGEHFIVDPKLSSTIVOLEKILKLEGLDYSLRQTTQTFQFDMTT 716
Db 1013 -----LSKAE-----TERLVKGIKERELEIKLLNEKNISLTKQIDQ----- 1048
QY 717 AGKALISK--SNLRQEQKQFLOEAQFFLGRV-----DLDKAIAKAEKAL 759
Db 1049 -----LSKDEVGKLTQIQKDL-EIQALHARISSTHTQDVVYLOQLOQAVAMEREKVF 1102
QY 760 -VTKKATKNGQLLERSINKAVLAYNNSAIKKANVRKLEKELDLTLGLVEGKGFLAQATMV 818
Db 1103 AVLNEKTRENSHLKTEYHKWM--DIVAAKEAALIKLQDENKKLSLTFESSGQDMFRET 1159
QY 819 QGVYLLKTPLEPYIIGLVNVPDKSGKLIYALDMSDTTIGEG-----QKDAY 865
Db 1160 QNLSRI-----IREKDIEIDALSQKQOTLLAVLQTSSTGNEAGVNSHOFEEILQBRDKL 1214
QY 866 GNPILNVD-----DNEGVHALAVATLADYEGLDIKTLNLSKLSQLT 907
Db 1215 KOQVKOMEWKOQVMTTVQNMQHESALQOELHLOAQVVLVSDN-----NSKLQ----- 1264
QY 908 SIRQVPTAAYHRAGIFQA-----IQNAAAEQQLLPKPGTHSEKSSSESANSKDRGL 960
Db 1265 -----VDYTGLIQSYEQNETKLKQFQELAQV-----QHSIQLCNTKDLLL 1306
QY 961 QSNPKTNRGHSAILPRTGS 980
Db 1307 -----GKLDIISPOLSS 1318

RESULT 72
US-09-949-016-7404
; Sequence 7404, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7404
; LENGTH: 2047
; TYPE: PRT
```

; ORGANISM: Human
US-09-949-016-7404

Query Match 3.2%; Score 164.5; DB 4; Length 2047;

Best Local Similarity 18.9%; Pred. No. 0.019;
Matches 193; Conservative 173; Mismatches 389; Indels 267; Gaps 46;

```
QY 29 KEPILKQTOASSISGADYAESGSKL-----KINETSGPVDOTVTDLSFKRTTPEKIK 84
Db 442 KERILAQASVVEVFLQOALSDAENIMRLSSLNQDLSLAEDNLK--LKMRIEVLKEX 499
QY 85 DNLAGPQRELKAVTENTESE---KQITSGSOLBSKESLSLNTKTPVPTSWEICDFIT 141
Db 500 SLLSQEKEBLQMSLLKANNEYEVIKSTATRDISLDSLHDLRLNL---EAKQELNQSIS 556
QY 142 KGNLTVGLSKSGVEKLSQ-----TDHLV-----LPSQADGTQLIQVASFATPDKKTAI 191
Db 557 EKETLI-----ABIEBLDRQNEATKHMILIKQOLSKQNEGDSIIISKLQDLNDEK--- 609
QY 192 AEYTSRAGENGESIQLDVDGKEIINEGEVFNYSYLLKQVITPTGYKHIGQDAFVDNKNIAE 251
Db 610 -----RVHQLEDDKMDITKELDVOKEKLIQ-----SEVALNDLHLTK 646
QY 252 VNLPSLETISY-----AFHALAKQIDLDPNLKA-IGELAFFDQNTGKLSLPRQLM 304
Db 647 QKLEKVENLVDQLNKSQESNVSIQKENLELKEHIRQNEEELSIRNE-----LMQSLN 700
QY 305 RLAEAFKSNHIKTIIEFRGNSLKVIGEASQNDLSQL-MLPDGLKIESEAFNGPDD 363
Db 701 QDSNSNFKDTLLKEREAEVRNLK-----QNLSELEQLNENLKKVADV-----K 744
QY 364 HYNRVVLWTK-----SGKNPSGLATENTVYVDPKSLWQSPED-----YTKW 407
Db 745 MENELVLVACEVVRHOLEECLAGNQLSL-EKNTIV---ETLMEKEIEAEELCWAKRL 800
QY 408 LEEDFTYQKNSVTFGSNKGLOKVRKNK--LEIPKQHGVTITEIGDNAPFRVDFQNTK 464
Db 801 LEEANKYK-TIEELSN-----ARNLNTSALQLEHEH-----LILK 835
QY 465 LKDYLEEVKLPSTIRKIGAFAPQSNLKSFEASDDLEIRKEGAFMNNRIETL--ELKOK 522
Db 836 NOKKMEIAELKNIEQMDTHKETQV-----LSSSEBEQKLTQLINKKEIFIEKLER 891
QY 523 LVTIGD-----AAPHINHIYAILVPESVOEIGRSAPFRQNGANNLIFMGSKVKTIGEMAP 576
Db 892 SSKLEELDKYSQALRKNEI-----LRQTEE-----KORSLSGME 928
QY 577 LSNRL-EHLDLSEQKLTTEIPVQAFSDNALKEVLLPAS---LKTIRE--EAFKKNHLKO 629
Db 929 ENNHLOEBELERLREESQRTAPVA--DPKTLDSVTELASEVSQLNTIKEHLEBEIKHHQKI 986
QY 630 LEVASALSHIAPNALDD--NDGDE-QFDNKVVVVKTHNSYALADGEHFIVDPKLSSTIV 586
Db 987 IEDQNSQWQLLSQLOEQKEMDFRYQHEQMATHYQTFLEKDEE-----IK 1034
QY 687 DLEKILKLEGLDYSLRQTTQTFQFDMTTACKALLSKSNLRSQGEKOKFLQBAOPFLGRV 746
Db 1035 SLQKTIEQIK-TQJLHEERQDIQTDNSDIFQETK--VQSLNIENGSEKHDLSKAE----- 1085
QY 747 DLDKAIAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKKANVRKLEK-ELDLTLGL 805
Db 1086 -----TERLVKG--IKERELEIKLNEKNISLTK-QIDQLSKDEVGKLTQI 1128
QY 806 VEGK-----GPLAQATMVQGVVLLKTPPL-----PEY 834
Db 1129 IQOKDLEIQALHARISSTHTQDVVYLOQLOQAVAMEREKVFVNLNEKTRENSHLKTEYH 1188
QY 835 IGLNVVFDKSGKLIYALD-----MSDTIGEGQKDAYGNPILN-----VDEBNEY 879
Db 1189 KQMDIYAAKBAALIKLQDENKKLSLTFESSGQDMFRETQNLRIREKDIIDALSQK 1248
QY 880 HALAVATLADYEGLDIKTLNLSKLSQTSIR-----QVPTAAYHRAGIFQATQNAAEAE 934
```

Db 1249 QTLAVLQTSSTGNEAGVNSNQPELLQERDKLKQVKMEWKQVMTTVQNMQHESA 1308
QY 935 QL 936
Db 1309 QL 1310

RESULT 73

US-09-710-279-3188
; Sequence 3188, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3188
; LENGTH: 1279
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-3188

Query Match 3.2%; Score 164; DB 4; Length 1279;
Best Local Similarity 18.0%; Pred. No. 0.01;
Matches 184; Conservative 187; Mismatches 363; Indels 286; Gaps 48;
QY 20 HNEVFSLVKBPILKQTOAS8SISGADYAESGSKKLINETSQVDDTVTDLFSDEKTT 79
Db 244 YNEQNTINTPTDDEEKQKALDKLTKAOGYKYNQVDAQTNQVSDAKTEAID----T 299
QY 80 PEKIDNLAGPREQ-ELKAVNTESKQITGSQLQSKESLS-LNKTVPSTSNWIEIC 137
Db 300 ITNIQANVAKPSARVELDSKPEDLKRQTNATPNATEBEKQDAIQRLNGKRDVKN--LI 357
QY 138 DFITKGNLVLGSKGVEKLSQTHLVLPSQAADGTQLIOVASFAFTPDKKTATAEYTSR 197
Db 358 NQDRDNEVEQHNTGLQEL-ETIH-ANPTRKSDALQELQTKFISQT----- 402
QY 198 AGENEISQLDVGDKEIINEGEVFNLSLLKVTIPTGYKHIGQDAFVDNKNIAEYNLPES 257
Db 403 -----ELINNNDATNEEKDEAKRLLE-----ISKNTIINQAOQN----- 440
QY 258 LETISDYAFALHAKQID-LPDNLKAIQELAFDNDQITGKLSLPQLMLAER---APKS 313
Db 441 -----NQVDNAKDN--GMNEIATIPATTIKTDAKTAIDKAEQVQVTTIING 484
QY 314 NHIKTIEPRGSLKVLIGEASFOQNDLSQLMLPDGLEKIESEAFTPGNPGDHYNNRVLWT 373
Db 485 NNDATDEKAEARKLVKA-----KIEAKSNITNSDTEREVN----- 521
QY 374 KSGKNPSGLATEN-----TYVNPDKSLWOESPE-----IDYTKWLEEDFTYQKNSVTGF 422
Db 522 --GAKTNGLEKINNTPSTQTKTNAKQINDKAQEQLIQINNTPATEE---EKQEAETR 576
QY 423 SNKGLOKVRNQLEIPKOHNGVTTITEIGDNFRNVDFQNTLRYDLEEVKLPTIRKI 482
Db 577 VNAGLAQAQIN-----NAHSQEVNESKTSI-----ATIKSVQPNVINKPTAINSL 625
QY 483 GAFQSNLKSFEASDD--LEBIEGAPMNNRIETLEKOKLVITGDAA--PHINHIYA 538
Db 626 ---TQEAANNQKTLINGDNATODEKEAA---KQVLTQKLINEQIQIHSTQDNQVDNKA 679
QY 539 IVLPSVOEIGRSAPFRQGANNLIFMGSKVKTGLGEMAFLSNRLHLDSQKQLTEIPVQ 598
Db 680 QAI-TAIKLINANAKRQDAINL-----TNLAESK---KSDIR 714

QY 599 AFSDNALKEVLLPASLKTIREEAFKKNHLKQLE--VASALSHI---AFNAL-DDN--DGD 650
Db 715 ANQDATTET-----KNTAIOISIDDTLQAOARNNINGANTNALVDENLEDGK 759
QY 651 EOPDNKVVVTHNSYALAD-----GEHFTVDPKLSSTIVDLEKILKLTIEGLDYSLR 704
Db 760 QKL-ORIVLSTQTKTQAKADIAQAGORSTIDQONATT-----EKKQEAERLNOETNG 814
QY 705 QTTOTQPRDMTTAGKALLSKNSLRQGEKQKFLQBAQFFLGRVDLDKAIKAEKA--LVTK 762
Db 815 VNDRIQ-----AALANQNVTD-EKNNILETIR-----NVEPIVIVKPKANEIIRK 858
QY 763 KATNGCOLLER-----SINKAVLAYNNSAIKKANVKELEKELD 800
Db 859 KAAEQTTLLINQONQATLEEKQIALGKLEVEKNEALNQVSAHSNDVK----- 906
QY 801 LLTGLVEGKGLAQATWVGVLKLPPLPEYVIGLN---VYFDKSGKLIYALDMSDT 856
Db 907 ----IVENNG-IAKISEVH-----PETIKRNAKQEIQDQAQSI-----DT 943
QY 857 IGEQKDAYGNPILNVDENEGYPHALAVATLADYEGLDIKTILNSKLSQL-----T 907
Db 944 INANKSTNEKSAADRNV-----VAKID-----AINNITNATTQLVNDAKNSGNT 991
QY 908 SIRQVPTAAVHRAGIFQAIQNAARAEQLLPKPGTHSEKSSSSSANSKORGLQSNPKTN 967
Db 992 SISQILPSTAVKTNALAAALASEAKKNNAIIDQ--TPNATAEKEEANNKVDRLQEEADAN 1049

RESULT 74

US-09-543-681A-5182
; Sequence 5182, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5182
; LENGTH: 1534
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5182

Query Match 3.2%; Score 163.5; DB 4; Length 1534;
Best Local Similarity 20.6%; Pred. No. 0.015;
Matches 237; Conservative 173; Mismatches 433; Indels 309; Gaps 61;
QY 35 QTOASSISGADYAESGSKKLINETSQVDDTVTDLPSDKRTTPEKID-NLA----- 88
Db 99 QGEPSSLIENOSATIKTEKGD-LIRTKLVNESITPLFEIKQEPPOSTSINIGTHPLH 157
QY 89 --KGPREQ-----ELKAVTENT-----ESEKQ-----ITSGS 113
Db 158 PYNSTPNFLIILLAEKLSDDKMPDILDKNNGVNVVERSLPEKKNKYVPSVIYSGK 217
QY 114 QLE-QSKESLSLNKTVSTSNWEICDFTIKGNTL-VGLSKSGVLEKLSQTD-----HLV 164
Db 218 NLYIQSNEFFNTQSDIFPATN-----DLIATGNNAKVVYVYRFG--HLAKWLDYSHDNAHF 270
QY 165 LPSQA---ADGTQLIOVASP-----AFTPDKKTATAEYTSRAGEINSEIQLDQVD 210
Db 271 YRDDVNKYNIDETKPKLHVHPTKIGSHVEFIVD---SASEYMKAGNN-----LVLD 322
QY 211 GKEINEGEVFNLSYLLKKVTIPTGYKHIGQ-DAFVDNKNIA--EVLNPESLETISDYAPA 267
Db 323 FKDSIALERKF-PFLEKEI---KKYRRISDPFETILAKNILLNANNLNLSLKKSSNLS 378

Db 1848 TS-VEFKDANGTGSSTKITKOGTITPAN-----GAGAAGANTANTISV--TKDGIS 1898
Qy 859 EGQDAYGNPILNVDEDEGHALAVATLADYE-----GLDKITILNSKLSQLT 907
Db 1899 AGNK-AVTNVSLGKKFGDG-HTLANGTVADPEKHYNAYKDL/TNLDKGDANN-----1950
Qy 908 SIROVPTAAYHRAGIFQALQNA--AAEAQQLPKPGTHSEKSSSESAN-----SKDRGLQ 961
Db 1951 -----PTVADNTAATVGLDRLGLGWISADKTGTGP--NOEYNAQVRNANEVFKPSGNGIN 2003
Qy 962 SNPKTNRG 969
Db 2004 VSGKTLNG 2011
RESULT 76
US-09-248-796A-19313
; Sequence 19313, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19313
; LENGTH: 1173
; TYPE: PRP
; ORGANISM: Candida albicans
; NAME/KEY: UNSURE
; LOCATION: (210)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-19313
Query Match 3.2%; Score 162.5; DB 4; Length 1173;
Best Local Similarity 19.7%; Pred. No. 0.011;
Matches 215; Conservative 176; Mismatches 390; Indels 313; Gaps 53;
Qy 20 HNOEVFLVKPEILKQTOASSSI-----SGADYAESGSKLK-----INET 61
Db 272 HKQEI-SRYKPELNKQQQESSLKQQLLEISSKQALYAKQSRFLKFNKXDRDSWLNT 330
Qy 62 SGPVDDTVTDLPSDKRTTPEKIKONLAKGPREQELKAVTENTESKQITSGQLBSKES 121
Db 331 ISKLKQITDKQDMRHHSNEYK-----TRESSLE--ELSESIKLNDLSNDEHIKT 381
Qy 122 LSLNKTVPSTNWEICDFITKGN-----TLVGLSKSGVEKLSQTDHLVLPQAADGTQ 174
Db 382 LANLKTINDSQQTQTLVDQKILWRDEIRLKSVDLSLNTDNLATNIV--NQTMRAQ 439
Qy 175 LIQVASPAFTPKKTAIAYTSRAGEN-----GEISQL-DVQDK-----EINNEGVSFNSY 224
Db 440 AQGIA-----AVKQIAQLNLSDRVYGTVAELFNVDNKYKTAARVIAGNSLPHIV 489
Qy 225 LLKKVTIPTGYKHI-----GQDAFVDNKNI--AEVNLPESETISDYAFALHAKQIDLP 277
Db 490 VDTVDVTAATIMEELIRNKAGRTVTFPLNRIDNIEVYDPDSEN-----QCLFLI 538
Qy 278 DNLKALGELAPFDNOITGKLSLPROLMRLAE--RAPKSNHIKTIETFRGNSLKVIGEASFQ 335
Db 539 KKLKNEQYKAINQIFKTLVVSLLKGGELSRVK-----575
Qy 336 DNDLSQMLPDGLEKIESEAFNGPDGDDHNNRV-VLMTKSGKNPSGLATENTYVNPDKS 394
Db 576 ---LSCITL-DG-DRVDTRGVLSGGYRDYKNSRIDALKIOTRKQE---LEKT-----DRE 623

Qy 395 LMQSPSIDYTKWLEEDFTYQKNSVTGFSNGKLG-KVKENKNLEIPKOHNGVTITEIGDN 453
Db 624 LVKVTETEST-----NSQLKNLNNELQNVRLDRLSVSKPEPIKIEUSQL---669
Qy 454 AFRNVDFONTKRYDLEEVKLPSTIRKIGAFAPQSNLNKSFASDDLEETKEGAFMNNR 513
Db 670 -----TNKFNLDQ-----EISLKS--NLQNLQNTKNSIKVNLK 702
Qy 514 IETLEKDLVTIGDAAFHINHIYVLPESVQIGRSAPFRQNGANNLI FMGSKVKTLGE 573
Db 703 QHELEL-----NSEFTQVLTQDEQ-----NEIDELSK 729
Qy 574 MAP-LSNRLEHLDLSEQQLTEIPQAFSDNALKEVL--LPASLKTIREAFKKNHLKQL 630
Db 730 LAIELESKLDHIVTRSSBELDTKI-----SGIESEVINNLQPKLNKRYQEQ-OKHOQOL 782
Qy 631 EV--ASALSHIAFNALDNDDEQFDNKKVVVKKTHNSYALADGHEFIVDPDKL--SSIV 686
Db 783 QLOQESSKSDTKSNLEVEELQOELENL-----HIQDTSQLNSQVVE 826
Qy 687 DLEKILKLEGLDYSTLRQTQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRV 746
Db 827 NLTKINBEINNC--QELAQAANKQKIKIKNIEKELKQTNLL--NQKLIKSM-----RD 878
Qy 747 DLDKAIK-----AEKALVTKKATK--NGOLLER--SINKAVLAY--NNSAIKKANVKRL 795
Db 879 DANOKIRELGVLPPEAFQSEKIDQYSSDQLLSKLNGINQELTKYSHINKKAIQFNLPNR 938
Qy 796 EKE-----LDLLTGLVEGKPLAQATWVGQVYLLKTPPLPPEYVIGLVN 839
Db 939 QKEDLMARRIDLNNKASIEENLITNLOQKNDKAIKSFQVAKSFQ-----985
Qy 840 YFDK-----SKLIYALDMSDTIGEGQDAYGNPILNVDEDEGHALAVATLADYEGLD 894
Db 986 IFEKLVPRGTGNLIMQKKNNNNNVATDDFNSD--NDDDDDDDD-----DIDNYSQVA 1036
Qy 895 IKTILNSKLSQLTSTIRQVPTAAYHRAGI--PQATQNA-----AAEA 933
Db 1037 ISVFSNKNDEQQRIEQLSGGQSKSLCALIALIFQNCDDPAPFLPFLDEIDSNLDYQYRISV 1096
Qy 934 BQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFVYGLGYTS 993
Db 1097 ARLLHIELSRNEDNNNE--GRSRGAQFICTTFR---PELLQLSGDK---FYGV---TF 1144
Qy 994 VALLSLITAIKKK 1007
Db 1145 SNKVSSVNEINKEE 1158
RESULT 77
US-08-790-912-3
; Sequence 3, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912

US-08-790-912-2

Query Match 3.2%; Score 162; DB 2; Length 2052;
Best Local Similarity 18.2%; Pred. No. 0.029;
Matches 208; Conservative 156; Mismatches 392; Indels 384; Gaps 51;

QY 30 EPILKQTAQSSGADVAESSGSKLKINETSQVDDTVT-----DLFSDKETTPEK 82
DB 608 QABESTTNSEKVSPTDTSNTGE-----VSNPDSSTTSVGBSNKPEHNDKENSEK 661

QY 83 IKDNLAKPREQELKAVTENTSEKQITSGSQESKESLSLTKT-----VP 129
DB 662 TVEEVPVNPNEGTVG-ISNQETEPVQPAETQNSGKIANENTGEVSNKPSDSKPPVE 720

QY 130 STSNWEICDFTK-----GNT-----LVGLSKSGVEKLSQTHLVLPQQAAG- 172
DB 721 ESNQPEKNGATAPKPSNGNTTSENQGTPEKKLELRNVSDIELYSQTNQYRQHVSLDG1 780

QY 173 -----TQLIOVASPAP-----TPDKK-----TATAEYTSRAGENGESQ--- 206
DB 781 PENTUTYFVKVSSAPKOVIPVASITEKRNQGVYKITAQKLOQELNENKYVDNFSF 840

QY 207 -LDVDGKEIINSGEVNSY--LLKKVT-IPTGYKHIG-----QDAFV 244
DB 841 YLDKAKE---ENTNFTSFLVKAINQNQPSCTYHLLAASLANEVELGPDERSYIKDTFT 897

QY 245 -----DNKNIAEVLNPSL-ETISDYAPAHALAKQIDLPNLKAIGELAFPDNQITGK 296
DB 898 GRLIGEKDGIYAIYNLKKPLPENLSGATVEXLSLKNV-----AISGK 940

QY 297 LSLPRLMLAERAFKSNHIKTIEFRGSLKLVIGEASFQDNDLSQLMLPDGLESAPF 356
DB 941 -----NDIGSLANEATNGYKIKQHVVDG---VLAGER-----GVGGLTAK 979

QY 357 TGNPGDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKMLEEDFTYQK 416
DB 980 QSSIAESSFKGRIV-----NTYETTD-----1000

QY 417 NSVTGFSNGLQKVRKNQKLEIPKQHNGVTITEIGNAPRNVDFONKTLRYDLEVKLP 476
DB 1001 -----AYNIGGLVGLHTGKNASIAKSKATVTIS---SNTNRS-----D 1035

QY 477 STIRKIGAFQFQSNLKSFEASDDLEIKE-----GAFMNRTE-----TLBLK 520
DB 1036 QTVGGLAGLVQDAHIQNSYABGDINNVAHFKQVAGVAGILMDRTSGEEKHAGELTNVLS 1095

QY 521 DKLVITGDA--AFH-----INHIYAIVLPSVQBIQSAFRQNG---ANNL 561
DB 1096 DVNVTNGNALTGYHYTGKMKVANTPSSKANRVENVTL-EKDEVVSKESFEERGTMLDASQ1 1154

QY 562 IPMGSKVTLGBMAFLNRLEHLDLSEQKQLTEIPVQAFSDN-ALK-----EVLPLASLKT 616
DB 1155 VSKAEIINPL-----TLPTVEPLSTSGKSDSFKIAHYQANRALVYKNIKLLPFYNS 1209

QY 617 -----IREEAFKKNHLKOLEVASALSHIAFNALDNDGDQFQDNKVYVKTTHNSYA- 667
DB 1210 TIVKYNLVKENS-----LYQKELLSAVNMKDDQVITDIVSNKQFANKLL--HYNDHSS 1263

QY 668 -----LADGEHFTVDPDKLSSTTVDLEKILKIEGLDYST- 702
DB 1264 EKFDLKYQTDPAFLPEYNLGNLTGLYTPNQFLYDRDSI-----VKEVLPLOKLDYQSD 1317

QY 703 -LRQT-----TQTPRDMTMTAGKALLSKNSLRQGEKQKFLQEAOFFLGRVLDKAI 752
DB 1318 AIRKTLGISPEVKLTVELDQ-----FSKTKQNLGDSLKKLSADAGLAS---DNSV 1367

QY 753 AK--AEKALVTKATKQNG-QLLERSIN-----KAVLAYNNSAIKKANVKLEKLD 800
DB 1368 TRGYLVDKIKONKEALLGLTYLERWYFNPNYQVNVKOLVMYHPDFGKNGTSPDITLIE 1427

QY 801 LATGLVEGKPLAQATWQGVYLLKPLPLPEYIIGLVNYPFKSGKLIYALDMSDTIGE- 859
DB 1428 L-----GKS-----GFNLLAKNNVDYTGISLASQHGAT 1456

QY 860 ---CQKDAYGNPILNVDNEDNEGYHALAVATLADYEGL--DIKTILNSKLSQTSIROVPT 914
DB 1457 DLFSTLEHYRKFVLPNTSNNDFKSETKAYIVESKSTIEVKT-----KQGLA 1504

QY 915 AAYHRAGIFQAIQAAAEAEOL-----LPKPGTHSEKSSSESANSKDRGLQSNPKTNR 968
DB 1505 GTKSIGYVDRTSATWKYRNWVPLITLTPERSVFVISTWSSLGFAYDRVRSDDHKAGK 1564

RESULT 79
US-09-134-000C-5785
; Sequence 5785, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5785
; LENGTH: 1282
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5785

Query Match 3.2%; Score 161.5; DB 4; Length 1282;
Best Local Similarity 19.2%; Pred. No. 0.015;
Matches 211; Conservative 149; Mismatches 341; Indels 397; Gaps 54;

QY 147 VGLSKSGV-----EKLSTQ-----DHLVLPQAAADGTOLI 176
DB 68 VGVAQSGTYYVDCGEGNANDQSPASAWDPPEKYNQTEFQPGDHVLLNAQSTWNNQLL 127

QY 177 Q-----VASPAFTPDKKTAAIAYTSRAGENGESISQDLDVKGKIIINEGEVFNYSYL 225
DB 128 HPKNGTAAQKIVDFDVTNDKGTEIPETTR-----PIINGGGTSTGT 172

QY 226 LKKVTIPGYKHIGQDAFVDNKNIAEVLNPSLETISDY-----A 265
DB 173 FKRAI--SGAVOLVNOQYWDISNLEVTNTPB-LDNLEGYKKGPDQAQAGILVLGYEQNRT 229

QY 266 FAHLALAKQ-----IDLPN-----LKAIGEL-----AFPD---NOITGKLSLPRQ 302
DB 230 FNSVTIRNNYVDVQTEYILNLSGNTATKRLKAVGGIIVLGSWFDENGWVTA-ANDHRT 288

QY 303 LMRILAERAFKSNHIKTIEF-----RGNSL-KVIGEASFQDNDLSQLMLPDGLE 349
DB 289 TTGFNDILINNVIORVGLGIRTKADSDTSRGNTFYKTFSTNITIRNNYLEDI-AGDGI- 346

QY 350 KIESEAFQNGPDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQ----- 397
DB 347 -VLSEAKSGGVVEGNVAVRMCNADYGTQNYAGVWA-----MSVDDGLFQYNEVYGIKYGFN 401

QY 398 --ESPEID-----YTKMLEEDF-----YTKQNSV-----TGFSN- 424
DB 402 DAAYDVMQSNVYQNYSHNTTGGFLLMLSDQKNSVIRYNTSANDGGGNRRRTGKDP 461

QY 425 --KGLQKVK-----RNKNLEIPKQHN-----GVITTEIG-----DNAPRNVDPQ 461
DB 462 RGAGYNYKEQSIIFYWVKNDGAAMPHTHNNTTIVVGDISTSLFEGEGNSDSNGTVANFY 521

QY 462 NKTIRKVDLEVKLPSTIRKIGAFQFQSNLKSFEASDDLEIEKEGAFMNRRIETLELK- 520
DB 522 NNILYKSGTGQLKFLSN-----YPTNGTQPIERK--MYDNPEKYPKNNVWPKEJAT 571

QY 521 -----DKLVITIGDAAPHINHIYAIVLPSVQBIQSAFR-2--QNGANNLI FMSGKVKT 571

Db 572 EKSGATVEKLVSSGN-IFKQPQLIEITDNPKEVKELAEQBEFTTLKPTKONVVEFTSKER-- 628
Qy 572 GENAFSLNRLEHLDLSEQ-----KOLTEI--PVQAFSDNALKEVLLPASLK---TIREEA 621
Db 629 -----LRQRAQMFRLAKENSFAIGKGLSEVNSPAEDFFGNSLKNKVLIDIGAQAQASTIEKSI 683
Qy 622 FKQHLKQLEVASALS-----HIAFNALDDNDGQDFQDNKVVVTKHNSYALADGE 672
Db 684 RYQNVQ--LEISSATGVYVPLPQVELTY-----BEVVNEEVVATGKKEFKL----- 728
Qy 673 HFIVDPKLSSTIVLEKILKJEGLDYSLRTTOTQPRDMTITAGKALLSKSNLROGEK 732
Db 729 QWEAIOEKINTAGTVE--VAATVIGLPIDAVKVTAKVSF-----EGELGEGKD 775
Qy 733 QKFLOBAQ--FFLGRVLDK---AIAKAEXALVTKKATK-----NGOLLERSINKAVLA 781
Db 776 TVKLKTAQYAYVQKSDGNRAYSAGTAAISSGDAYKYPYGVNYTGNVYALKKNASSAG 835
Qy 782 YNN-----SAIKANVRLEKELDLTLGLVEGKGFLAQAATWQ--GVYLLKT 826
Db 836 YNRRIVVEIDTQELKKNYQSLKSANLELVNMYDAMNAGNTNDRKNTQFQVDVYGTDT 895
Qy 827 -----PLPLPEYV-----IGLVNVPDK---SQ----- 845
Db 896 NMSNTITWNGPNLNVNEEFIAEQSFNTSSIMNQNTISIDISNYLRKLIQSGEKIP 955
Qy 846 -KLIVALDMSDTIGEGQKDAYGNPILNVDENEGYHA----- 881
Db 956 AKLSFLLAITDSRLPG-----YSDNAGFPAFSKEGAKAYQDFLTKLTLPTG 1004
Qy 882 --LAVATLA-----DYGLDIKTLNS--KLSQTSI-----RQV 912
Db 1005 QQLAEDSLAPKIVLSNVFVKHESIEVTEAGQAPKLPKTTIFYSDGSGOREVTNVNMSV 1064
Qy 913 PTAAYHRAGIFQAIQAA 930
Db 1065 PASSYQKEGIFTVVGRAA 1082

RESULT 80

US-08-956-171E-5235
; Sequence 5235, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 5235:
SEQUENCE CHARACTERISTICS:
LENGTH: 886 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5235:
US-08-956-171E-5235

Query Match 3.2%; Score 161; DB 4; Length 886;

Best Local Similarity 19.0%; Pred. No. 0.0097;
Matches 201; Conservative 149; Mismatches 348; Indels 360; Gaps 51;

Qy 2 KKHAKTVALTLTVSVVTHNQEVSLVKEPILKQTOASSISGADYAESSGSKLKINE- 60
Db 30 KSTLGVASVIVSTLFLITSQH-----QQAQAEINTTSD-----KISEN 67
Qy 61 -----TSGPVDDT-----VTDLFSDKRTTP---EKIKONLAK-----GPREQ-E 95
Db 68 QNNNATTTPPKDQNTQTPATQPAANTAKNYPAADESLSKDAIKDPALENKEHDIGPREQV 127
Qy 96 LKAVTENTESEK-----QITSGQLSQSKESLSLNTKVPSTSNWEICDFTKGNLTVLGSK 151
Db 128 FOLLKNNETQYHFFSFKDPADVYVTKKAEVELDINTASTWK-----KPEVYENNQ 181
Qy 152 SGVEKLSQT-----DHLVLPQAAQDGTOLIQVASFATPDPKKTAIAEYTSRAGENGESQ 207
Db 182 LPVRLVSYPVEDHAYIRFPVSDGTQELKIVSSQTQIDGGEFTNYDTKL----- 231
Qy 208 DVDGKEIINEGEVFNSSYLLKVTIPTGYKHIGQDAFVNKNIAEVLNLPESLETISDYAFA 267
Db 232 -VFAPKIYND-----PSLVKSDTNDVAVTNDQSSSVASNGTNTNSN--- 272
Qy 268 HIALKQIDLPDLNKAIGELAFPDNQITGKLSLPRQ--LMRLAERAFKSNHIKTIPEFGNS 325
Db 273 -----QNIISTINN-ANNQPAATTNMSQPAQPKSSTNADQA--SSQPAHETNSNGNT 320
Qy 326 LKVICEASFQNDLSQMLP-----DGLEK---IESEAFTCN--PGD----- 362
Db 321 NDKTNESNQ--SDVNQOYPPADESLQDAIKNPAIIDKHEHTADNRWPIDFQMKNDKGERQF 379
Qy 363 DHYNNRV-----VLWTKSGK-NPSGLATENTY-----VNPDKSLWQESPEIDYTKWLEEDF 412
Db 380 YHYASTVEPATVIFTGTPIELGLKTAFTWKKEVYEGDKKLVELSVYSDSK-----DY 435
Qy 413 TYQKNSVTGFSNKGLOKVRKNKLEIPKQHNGVTITTEIGDANFRVNDFQNKTLRKLYDLEE 472
Db 436 AYIREPPV-----SNGTREVKI-----VSSIEYGENIHEDYD-----YTLMV 471
Qy 473 VKLPSTIRKIGAFAPQSNLKSFEASDDELEIKEGAFMNNRIETLELKDCLVTIGDAAPH 532
Db 472 FAQPIT-----NN-----PDDYVDE-----ETYNLQKLL-----APYH 499
Qy 533 INHIVAILVLPESVQIGRSAPFGQANNLIFMGSKVKTLGEMAFSLNRLEHLDLSEOKL 592
Db 500 -----KAKTL-----ERQV 508
Qy 593 TEIPVQAFSDNALKEVLLPASLKTIREAPFKNHLKQL-----EVASALSHIAFNA 643
Db 509 YEL-----EKJQEKL-----PEKYAEYKKLDDQTRVELADQVKSATFE-NV 551
Qy 644 LDDND--GDEQPDNKKVVKTHNSVALADGGEHFIVDPDKLSSTIVDLKILKLEGLDYS 701
Db 552 TPTNDQLDQEAHFVWFSEENSESVMDG--FVSEHPFYTAT-----LNGQYV 598

APPLICANT: Kupke, Thomas
APPLICANT: Gotz, Friedrich
APPLICANT: Kempter, Christoph
APPLICANT: Jung, Gunther
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/645,193B
FILING DATE: 13-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.1540000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 990 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-645-193B-15

Query Match 3.2%; Score 161; DB 2; Length 990;
Best Local Similarity 19.6%; Pred. No. 0.011;
Matches 186; Conservative 152; Mismatches 345; Indels 268; Gaps 48;
QY 28 VKPEILKQTAQSSS-----ISGADYAESGSKLKLINETSIPVDDTVD----- 71
DB 77 VKESLLKYLIRSTRPYGMLSGVALGEFSENNIKIDKDSFHKKVDKIDQWLYKLVH 136
QY 72 -----LFSDKRTT-PEKIDNLAKGPREQELKAVTENTE 104
DB 137 YLESDDYTYKDSFVIWNOQNYIYNRLYLDNNSITENKRDVLSVKYNSILVFITHEN-- 194
QY 105 SEKQITSGS--OLEQSKESLS-----LNKTV-----PSTSNWEICDPTKG- 143
DB 195 SKKNITYEELVQLISSKYSIENKEEVKVFQBELINKKEIIFSDLRPTLENKPLDYIINSL 254
QY 144 ---NTLAVGLSKSGVEKLSQTDHLVLPQAADGTQLIQVASFAPTDKTAIAEYTSRAGE 200
DB 255 NPKNSLVG-----TLINISN-----EITKYSKMPLG 280
QY 201 NGDISOLDVDGKEINEGEVFNLSYLLKVTIPTGYKHIGQDAFVD-NKNIAEVLNIPESL- 258
DB 281 KGEYKYLD-----IVN-----LMSQLFVSKNVLQI--DTVIDYSRN-----ELKQSLA 321
QY 259 EYISDYAPA-----HLALKQI-----DLPDNLKAIGELAFDQNTQKGL 297
DB 322 DNISEAAYITLWLLSPNPHFQTKIRNYHEFFMDKYGEQVLVNLKQLLSIDNGFGYPKDQSY 381
QY 298 SLPLRMLAER---AFKSN-HIKTIEPRGNSLKVIGEASFQNDLSQMLPDGLEKTES 353
DB 382 SFSNNIAFLKEKYLLAIQNNSHIEITE---NDVKLE-----KNNTVSKINAPVSTE-IYS 433
QY 354 EAFTEGNGPDHNNRV---VLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWL-E 409

DB 434 EYFGNSIKGYEDFAVISPILGSNAGATGFRFTGNFIKKKNLQKQKEIVH-HYNNYME 492
QY 410 EDPYQKNSVTGFSNKGLOKVKRK-----NLETPKHONGVTITEIG----- 451
DB 493 NDLEISQLEAPLNSRNVNILNNNRIYNTCLNINLPKSIDINDIFIGATFNKLYLSK 552
QY 452 -----DNAFRNVDQFQ--KTLRKYDLEEVKLPSTIRKIG-----AFAFOSNNL 492
DB 553 HDSRIVFVSNMFNYEFGSELYKFLREISFEKTKFIQPIESGIDSLPFCPRIIYKNIL 612
QY 493 KSFASDDLEETKEGAFMNNRIETLE-----LKOKLVITGDAAFHNHI---YAVLPS 544
DB 613 KPATWKINSEMESETENMLNREPATIREKWHIPKDVIIAFGDNRLNLLNLDKHLIILKE 672
QY 545 VOEIGRSAPRQ---NGANN-----LJFMGSKVKTIGEMAF--SNRLEHL-DLSEQKQL- 592
DB 673 LKHGRIRILESFINESNNRMLEIVTPLYKTSLEQSFIIPKNRNHNFNKLKOWFSIH 732
QY 593 TEIPVQAFSDNALKEVLLP-----ASLKTIREBAFKKNHLKQLEVASALSH 638
DB 733 LSP- KTYQDNFIQDYLPPFITELKVNNFINKFFYKPEDEDFIK--LRLREDEDSQ 789
QY 639 IAFNALDND-----GDEQFNDKV-----KTHNSYALADGEHFIVDDPKLSSTIVDLE 689
DB 790 IYSFIKNWKDYCLNSELYDYSIVDVPVRYGPHVIEDIENPFMYDSLISINIIQSE 849
QY 690 -KI-----LKLISGLDYSLRQTTQTF-----RDMTAG-----KALLSKSNLRQGE 731
DB 850 FKIPKEFIVAISIDFLDYLEINKSEKEILINNAEDLYRSNDIREYKNLLAKLTNPKND 909
QY 732 ---KQKFLQEAQFFLGRVLDLKAJAKA-EKALVTKKATQNGQLLERSINK 777
DB 910 YEILKKEFPNLHFLFNKISILENLKTLQKLSLYTSRSLGFSFIHMRCNR 960
RESULT 83
US-09-248-796A-15026
; Sequence 15026, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15026
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-15026

Query Match 3.2%; Score 161; DB 4; Length 1132;
Best Local Similarity 20.3%; Pred. No. 0.014;
Matches 201; Conservative 149; Mismatches 369; Indels 272; Gaps 47;
QY 10 LTLTVSVVTHNOEVSLVKEPILKQTAQSSISGADYAESGSKLKLINETSIPVDDTV 69
DB 63 MTLSSGTSKPSVEESKSLKPTIHKT-----STDYLSK---AKTKAKBEKVITIEKSD 113
QY 70 TDLFSDKRTT-----PEKI-----KONLAKGPREQELKAVTENTESEKQITSGSL----- 115
DB 114 KTNSEERKTEPIQSEQLLTDKKNKVEPNSEVNLK---DNTDDMK-ATACALGPDKN 169
QY 116 -----EQSKESLSLNKTVPSTS-----NWEICDF-----ITKGNL 146
DB 170 TKDNDSKSETTQPKLARSESFADTSLSPVNESDTDFNENELAEIPEAKDGSVVAANVL 229

QY 810 GPLAQTMTQGVVLLKTPPLPYIYGL-----NVTYDKS-----GKL 847
Db 821 EELLEALLKANKETQVODLQKEIKALKEEIGNVQLEKAQQLSITSKVQELQNLKKGKE 880
QY 848 IVALDMSDTIGEGOKDA--YGNPILNVDEDNCGY--HALAVA-----TLADYEGLDI 895
Db 881 EQWNTWKAVLEKEKELANTGKWLQDBENESLKAHVQEVQAHNLEKEASSAQFEELEI 940
QY 896 -----KTIILNSKLSQITS-----IROVPTAAHYHRAGIFQA 925
Db 941 VLKKEKENELKLEAMLEKRESLSSKTQLLDVQDENKLFKQJIEQLKQOQYQQASSF-- 998
QY 926 IQNAAAEAOQLPKPOTHSEKSSS-----SESANSKORGLOSQNPKNRGRHSAL- 974
Db 999 -----PPHEELL-KVISEREKEISGLWNLDSLKD-AVEHQKKNRQOQVEAVELEAK 1051
QY 975 -----LPRTGSKGSFYVG 987
Db 1052 EVLKKLFPKVPSPNSLYGEWLHG 1075

RESULT 85
US-09-081-149-8
; Sequence 8, Application US/09081149A
; Patent No. 6506889
; GENERAL INFORMATION:
; APPLICANT: Sieburth, Derek
; TITLE OF INVENTION: RAS SUPPRESSOR SUR-8
; FILE REFERENCE: UTC-02938
; CURRENT APPLICATION NUMBER: US/09/081,149A
; CURRENT FILING DATE: 1998-05-19
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; TYPE: PRT
; LENGTH: 582
; ORGANISM: Homo sapiens
US-09-081-149-8

Query Match 3.2%; Score 160.5; DB 4; Length 582;
Best Local Similarity 23.6%; Pred. No. 0.0056;
Matches 135; Conservative 69; Mismatches 211; Indels 157; Gaps 30;

QY 113 SQLEQSKESLSLNTKTPVSTSNWEICDPIFKGNTLVGLSGVGEKLSQDHLVLPQADG 172
Db 3 SLSGKEKSKEDPKVPKSAKE-----KEAKSGGFGKSEKEPEKTK-----GKDAKDG 53
QY 173 TQLIQVA--SFAFTPD---KKTAAIAYTSRAGENGESIQDVGKEIINEGVFNLSYL--- 225
Db 54 KXDSSAAQPGVAFSVDTIKRPNPAPGTRKSSNAEVI-----KE-LNKCRENSMRLD 106
QY 226 LKXVTIPTGYKHIGQDAFVDNKNIAEAVNIPESLETISDYAPAHALAKQI-DLPDN---LK 281
Db 107 LSKRSI-----HI-----LPSSIKELTQLTLYLSXNKLQSLPAEYGCIV 146
QY 282 AIGELAFPDNQITGKLSLQRLMLAERAFKSNHKTIEFRGNSLKVGEASFPQDNDLSQ 341
Db 147 NLMTALSNSLT---SLPDSLNLKK-----LRMLDLRNKUREIPSVVYRLDLSLT 196
QY 342 LMLPDGLEKIESEAFTEGNPDGDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPE 401
Db 197 LVU-----RFRNRTTVEKOIKNLSKLSMLSIRENKIKQLPAIGE 236
QY 402 IDYTKWLEEDFYQKNSVTGPNKGLQKVRKNKLEIPKQHNGVITEIGD-NAPRNVDF 460
Db 237 L-----CNLIT-----LDVAHQLEHLPR-----EIGNCTQITNLDL 268
QY 461 QNKTLRKYDLEEKLPSTTKIGAPA---FQSNL-----KSFESDDEEIKEGAFMNR 513
Db 269 QHNEL-----LDLPDPTIGNUSSRLGLRYNLSAIPRSLAKSAALELN---LENNN 318
QY 514 IETL--ELKDKLVITIGDAAFIN--HIYAIVLPEVQIEGRSAFRQNGANNLIF-MGSKV 568

RESULT 86

US-09-107-532A-5007
; Sequence 5007, Application US/09107532A
; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5007:

SEQUENCE CHARACTERISTICS:

LENGTH: 956 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...956

SEQUENCE DESCRIPTION: SEQ ID NO: 5007:

US-09-107-532A-5007

Query Match

3.2%; Score 160.5; DB 4; Length 956;

Best Local Similarity 19.6%; Pred. No. 0.012;

Matches 179; Conservative 135; Mismatches 288; Indels 313; Gaps 47;

QY 18 VTHNVEVSLVKEPIILKQTAQSSSISGADYAESCK-----SKKINETSQPV 65

Db 188 ISHYQDKYIIDRTV-----NTVVG---ASRSKGFTFVPLDLNLSRAKI-OSNMVV 237

QY 66 DDTVTDLFSKRRTPPEKIKDNLAKPRBQELKAVTENTESEKQITSGSQLEQSKESLSLN 125

Db 238 NDPKGELFSASKTELEK-----RGYRVLV-----NIDDPLEMSFN 274
Qy 126 K---TVPSTSNWEICDFTKGNLTVG--LSKSGVEKLSQTDHVLVPSQAADGTQLIQVASF 181
Db 275 PLQLVIDSWANKDVHEASKRANTLTSMLFASQMG--TDNEFFYKSAKSAVNAIILITIVEH 332
Qy 182 AFTPD--KKTAAEYTSRAGENGEISQLD--VDGKEIINEGEVFNPSYLLKKVTIPTGYK-- 236
Db 333 CFNNDICIEKTIWYNVAQMLNELGSLFYTPDKTEKKNALDEVFN-----TLPOGNAK 385
Qy 237 -HIGQAFVDNKNIAEVLNPSLETSIDYAFAPHLAKQIDLPDLNKAIGELAFDQNTG 295
Db 386 IQYGSTSPAGKANGSI-----LSTASQ-----GIEMFTSDLF 419
Qy 296 KLSLPRQLMRLAERAFKHNHKTIFRONSUKVGE-----ASF--QDNDLSQIM----- 343
Db 420 KLT-SKMSIDLKEIGFP-----KSIQKVNT-NLVGKRISISFLRKENDVIRLKKYRVK 473
Qy 344 -----LPDG-----LKKISEA-----FTGNFGDDHNNRVVLWTKSGKNP 379
Db 474 KALGCVLNFNEYLOQDMDIIRYEEKSRAWRIEFPKQKQNSHDTSLVTKKFSKQNY 533
Qy 380 SGLATEN---TYVNDKSLWOESPEIDYTKWLEEDFTYQKNSVTGFSNKGQKVRKNL 436
Db 534 SDLEISTLRKXYTDQPTAVFWIPDYD-----FTGNFGDDHNNRVVLWTKSGKNP 560
Qy 437 EIPQHNQVTI-----TEIGNAFRNVDPNQTKRYDLEBEVKLPSTIRKIGAPAFOSN 490
Db 561 --PSNHVLASIFISQLYTELASNCKQTPD--KKCFRRVH-----FLLD 599
Qy 491 NLKSFESDDEEIKEGAFMNNRIETLEK--DKLVTIGDAEFHINHIYAIVLPESVQEI 548
Db 600 BFGNPAIDNMDGIMTVCLGRNMLPDLVIQSYQLETRYDKAF----- 642
Qy 549 GRSAPRQNGANNLIPMGSKVTALGEMA-----FLSNRLEHLD-----LSEOKQLTE 594
Db 643 --KITKENCQHILMSNDBETIEBISKKGKHTTINRSASSKHLDTDSSVTSSABQERV 700
Qy 595 IPVQAFSNALKEVLLPASLTKTIREAPKGNHLKOLEVASALSHAFNALDDN----- 647
Db 701 ITVERASQLIEGEQII--LRNLRHQDTKRKNIRPPI-----FNTKETNMPRYQF 749
Qy 648 ---QDGEQFD-KNVVVKTHNSYALADGE---HEIVD-----PD 679
Db 750 LAEDPTQDINEIDIAEHNLSQDQNOIPIYARFIKOLKTRLEYSIINNIPISEEDYQD 809
Qy 680 KL-----SSTIVDLKILIEGLDYSTLRQTTQTFQDRMTTAGKALLS---KSNLRQGEK 732
Db 810 YLNLGSGQQTINEEBLLKLV-----NTEVKSQDM--EGQVLSQEMKINIYLKES 857
Qy 733 QKFLQBAQFFLGRVLDLKA-----AKASKALVTKATQNGQLE-----RSINKAVLAVNN 784
Db 858 QK--EA-----IELVKVINGRLADENSADWVIK--TINSQILDTPFTQSLINDSIATHN 907
Qy 785 SAIKKANVKRLEKEL 799
Db 908 VQESKQIQMILEQQ 922

RESULT 87

US-07-853-913-2

; Sequence 2, Application US/07853913

; Patent No. 5338839

; GENERAL INFORMATION:

; APPLICANT: McKay, Ronald D.G.

; APPLICANT: Lendahl, Urban

; TITLE OF INVENTION: Nestin Expression As An Indicator of

; TITLE OF INVENTION: Neuroepithelial Tumors

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.

; STREET: Two Militia Drive

; CITY: Lexington

; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853,913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1805 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-853-913-2

Query Match 3.2%; Score 160.5; DB 1; Length 1805;

Best Local Similarity 18.0%; Pred. No. 0.03;

Matches 180; Conservative 164; Mismatches 372; Indels 283; Gaps 42;

Qy 102 NTESEKQITSGSQLE-OSKESLSLNTKVPSTSNWEICDFTKGNLTVGLSKSGVEKLSQT 160
Db 402 NAEVRAEVLPSLQTOAPEPLMLKATVPSS----- 433
Qy 161 DHLVPSQAADGTQLIQVASFPTDKKTAIA-----EYTSRAGENGEISQLDVGKEII 215
Db 434 --AILPELEBPGK--QQGHP---PDDLTSLATNLNPHHPPTLEAKDGSESSRSVSIQOE 486
Qy 216 NEGEVFNYSLLKK-----VTIPTGYKHIGQAFVDNKNIAEVLNPSLETISDYAPHLA 270
Db 487 DEGOIWE--LVEKEADIEVKVENSQAOKTOSGLDTEETQDSOGPQKET----- 534
Qy 271 LKQIDLPDLNKAIGELAFDQNOI-----TGKLSLPRQLMRLAERAFK 312
Db 535 -----LKAIGEEPLMSLKIQNYETAGKENCNSSTEGHLGLEGEKQIPLKLSLE 585
Qy 313 SNHIKTIEFRGNLSKVIGEAQFQDN---DLSQLMPLDGLKIESEAFNGNFGDDHNNRV 369
Db 586 EKNVESEKTLNENGVPVLSELLGKEDTRTEQDELMSPKGTLK----- 626
Qy 370 VLWTKSGHNPGLATENTYNNPDKSLWOESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOK 429
Db 627 -RPSLSGKE-----SQEVVRPSKEGNLES-----WTAFKE--SQHPLGFFGAEDQMLER 673
Qy 430 VKRKNLEIPKQHNQVITTEIGDNAFRNVDPQ-----KTURKYDLEBEVK 474
Db 674 LVEKEQDSFPKSP-----BEDQACRPLQKNEQEPLOYEABEQILRLIEKESQESIR 728
Qy 475 LPSTIRKIGAFAPQSNLKS---EASDDEEIKEGAFMNNRIETLEKDLKVLITIGDAAF 531
Db 729 SPEEDQEAQSLQKNEQEPLOYEABEQMLE-----RLIEKESQESLSK----- 773

QY 532 HINHIYAI VLPESVQIGRSAPRONGANNLIFMGSKVKTLGEMAFLSNR-LEHLDLSEOK 590
Db 774 -----PEENORIGPLEREN-QKSRLYLEENQETFPVPLESRNQRPPLSLVEEBE 822
QY 591 QLTETIPQAFSPNALKEVLLPDLAKTIR-----EFAFKNHLKQLEVASALSHAFNALDD 646
Db 823 ORIVKPLEKVSODSLGS-LAEENVQPLRYLEEDDCINKSLLED-----KTHKSLGSLD 875
QY 647 NDGD-----EFOFNKVVVKTTHNSYALADGCHFFIVDPDKLSSTIVDLE 689
Db 876 RIGDSIIIOQESQVSLRPPEDORIV-----NHLEKESQEFSSSEEBE 922
QY 690 KIL-KLIEGLDYSTLR-----OTTQTF-RDMTTAGKAL-----LKSNI----- 727
Db 923 QVMERSLEGENHESLSSVEKEDQMVESQLEKESQDSGKSLEDBESQTFGPLEKENAESLR 982
QY 728 -----RGEKQKFLQEAQFPLGRVLDL-K-AIAKAKA-LVTXKATKNGQLLERSINKAVLA 781
Db 983 SLAQDQEEQKLEQETQOTLRVGNQMAVSPPEKVDPELPKPLGNDQBIARSLSGKE--- 1039
QY 782 YNNSAIKANKVLEKELDLTLGLVEGKGLAQATWVGVLKLTPLPEYYIGLVNYP 841
Db 1040 -NOESLVSLEKGIETVKSLEIIE-----PLETABEDLERKSI 1079
QY 842 DKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDEDNQYHALAVATLADYEGLDIKTILNS 901
Db 1080 DTQEPL-WSTEVARTEVPEDEPPGSLGSVDNRE-----TLTSLKESQELSSL--G 1130
QY 902 KLSQTSIRQVTPAAYHRAGI-FOAJONAAAEQQLPKPGTHS-----EKSSSESAN 954
Db 1131 KWNVETRVDSQOCLQVEGLQEQHESLREVKQELPSSGNOQRWEDVVEGKAVQEQAP 1190
QY 955 SKDRGLQSNPKTN---RGRHSAILPRTGSGSFVYGILG 990
Db 1191 LATTGVGTEDKAEHLRGQGGE---EBAABEGELLQDIVG 1227

RESULT 88
US-09-713-273A-20
; Sequence 20, Application US/09713273A
; Patent No. 6620987
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Butler, Karla
; TITLE OF INVENTION: STARCH R1 PHOSPHORYLATION PROTEINS
; FILE REFERENCE: BBI158 US CIP
; CURRENT APPLICATION NUMBER: US/09/713,273A
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: 60/081,143
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: PCT/US99/07639
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 09/679,933
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 1493
; TYPE: PRT
; ORGANISM: Glycine max
US-09-713-273A-20

Query Match 3.1%; Score 160; DB 4; Length 1493;
Best Local Similarity 19.5%; Pred. No. 0.025;
Matches 218; Conservative 164; Mismatches 358; Indels 376; Gaps 60;

QY 26 SLVKEPILKQTOA-----SSISGADYAESGSKLKINETSGPVDVTVDLPSDKRTTPE 81
Db 38 SIHFQTVLCQTVABHQBSVLSVANSKGNLFLAPT-----FRGRLCVR 87
QY 82 KIKDNLAKG-----PREQELKAVTENTESEKQITSGSQLEFSQKE-SLSLNKTVPSTS 132

RESULT 89

US-09-914-259-25


```
;
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4883:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1967 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1967
; SEQUENCE DESCRIPTION: SEQ ID NO: 4883:
US-09-107-433-4883

Query Match          3.1%; Score 159; DB 4; Length 1967;
Best Local Similarity 19.6%; Pred. No. 0.044;
Matches 245; Conservative 174; Mismatches 386; Indels 444; Gaps 64;

QY 29 KEPILK-----QTOASSISGADYABSSGKS-----KUKINETSPV 65
DB 407 KEPVDKSELNQIDKASSVSPDYSYNALGPVLETAKGAYASEPVKQPEVNSETNKL 466
QY 66 DDTVDLPSDKRTTEPKIK-----DNLAKGRE---QELKAVTENTES-----EKQI 109
DB 467 KTAIDALNVDSKELQEVRAEQOQADYSAKTWREFKIAELQAKBEINNQTTPPKQSEI 526
QY 110 TSGSG--LEQSKESLSLNTV-----PSTSN-----WEI-----136
DB 527 DAATKALQDALQALAVDTVLQNAINTANSKREBEYTAQTWKALEDALTAVNPVNEDETA 586
QY 137 ----CDFITKG-----NTLVGLSGVGEKLSQTDHLVLPQQAADGTQLIQVASFAPTPDK 187
DB 587 TQSKVDEATRNLLEEAINNLVLLTEKPVLTFTETDKKALEREV-----VAKYSLENQN 638
QY 188 KTAIAEYTS--RAGENGISQLDVQKEIINEG-----EVFNSYL-----225
DB 639 KTKIKSITATUKKGET--VSTVELIGDDVTNETTITSAPKNLEYKYEYTLSTMTVMVDRGDG 697
QY 226 -----LKKVTIPTGYKHIGQDAFVDNKNIAEVLNLPESLETI--SDYAFAPHL 269
DB 698 DVTIELDNQIQLDLKKVEL---KNIKRTDLIKYENGKEIN--ESLITTVDDKRNYYL 751
QY 270 ALKQIDLPDLKAIQEL-----AFFDNQITGKLSLPLQMLRAELAEAFKSNH 315
DB 752 KITSKNQKTTLLAVKNIBETTVNGTPVKKVTAIADNLVS-----RTADNKFEEY 801
QY 316 IKTIIE-----FRGN---SLKVIGEASPDND---LSQLM-----LPDGLKIESEAFNGN 359
DB 802 VHYIEKPKVHEDNVYTNFKELVEAIONDPSKEYRLQGSMSARNVYPNGKSYITKE--FTGK 860
QY 360 -----PGDDHYNN 367
DB 861 LLSBSGKQFAITELEHPLFNVTNATINNVPENVEIERSGQDNFIASLANTMKGSVITN 920
QY 368 RVVLWTGSGKN-----PSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNS 418
DB 921 VKITGTLGRNVAGFVNMNDGTRIEN--VAFFGKLHSTSGNGSHT-----GG 967
QY 419 VTGFSNKGKLOKV-----KRKNLEIPQHNGVITIE-IGDNAFRNVDFQNKTLRK 467
```

```
DB 968 IAGTNYRGIVRKAYVDATITGNKTRASLLVPKVDYGLTDLHLIGTKALLT---ESVVVGK 1024
QY 468 YDLBEVLKPTSTIRKIGAPAFQSNLNKSFASDDLEIEIKEG--AFMNNRIETLELKDKLVT 525
DB 1025 IDV-----SNPVEVGAIASKTWPVGTVSNVSVAKIIRGEBELFGSNDVDDSDY-----1072
QY 526 IGDAAFHINHIYAIVLPESVQETRSAPRQNGANNLIF---MGSKVKTGLGEMAFSLNPLE 582
DB 1073 ---ASAHIKOLYAVEGYSS---GNRSFRKSKTFTKLTKSQADAKVTFN---ITADKLE 1122
QY 583 HLDLSEQOLTEIPVQAFS-----DNALK--EVLPLASLXTI-----REEAFKKNHL 627
DB 1123 S-DLSPLAKLNE--EKAYSSIQDYNAYNQAYKNLEKLIFFYNKYDIYVYQGNKLKHEHL 1179
QY 628 KQLEVASALSHIAFNALDNDGDQFD--NKVVV-----KTHNSVALADG-----EHF 674
DB 1180 NTKEVLSV-----TAMNNEFITNLDEANKIIVHYADGTKEYFNLSSSSSSGLSNVKEYT 1233
QY 675 IVD-----PDKLSSTIV-DLEKILKLEGLDYSTLRQTTQTFQFRDWTAGKALLSK 724
DB 1234 ITDLGIRKYPNIVQKDNMTTLVNDIKSILESVE-----LQSQTMVYQHLNRLGDYRVN- 1284
QY 725 SNLRQGEKQKFLQBAQFPLGRVLDLKAIAKAEKALVTKATKQNGOLLERSINKAVLAYNN 784
DB 1285 -----AIKDYLEES-----FTDV-----KENLNLITKLQVNEHQNLN-----DS 1320
QY 785 SAIKKANVYKLEK-ELDLTLTGLVEGKGPLAQATMVQGVYLLK-----TPPLPEYVI 835
DB 1321 PAARQMIRDKVEKNKAALLLGL-----TYLNRVYGVKFGDVNIKELMLFKPDFY- 1369
QY 836 GLNVYPKSGKLIYALDMSDTIGEGQK-----DAYGNPILNVDEDEGNYHALAVATL 887
DB 1370 -----GEKSVLDRLIEIGSKENNIKSGRTFDAGQ-----VL 1402
QY 888 ADY-EGLDIKTLNLSKLSQTSIRQVPTAAVHRAGIFQAIQNAABAEQLLPKQTHSEK 946
DB 1403 AKYTKSGNLDAFLNRYNQLFTNIDMNDWFDIDATEDHVYIAERASEVEEI-----1452
QY 947 SSSSESANSKDRGLQSNPKTNRGHRSAILPRTGSGSFVYGILOYTSA 995
DB 1453 -----KNSKHRAF-DNLKRSHLNT-ILPLLNDIKAHLYLISNYNAIA 1493

RESULT 93
US-09-949-016-9776
; Sequence 9776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9776
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-9776

Query Match          3.1%; Score 158.5; DB 4; Length 1055;
Best Local Similarity 19.2%; Pred. No. 0.019;
Matches 187; Conservative 163; Mismatches 303; Indels 321; Gaps 52;
```

QY 11 TLTVSVVTHNOEVFLSVKEPIKQOQASSISGADYAESSGSKLKINETSQPVDDTVT 70
Db 21 TLTVN-----FSEWQE-----REETSQVLSALHFF-----PHLIQNEVS-RLIEDIQ 63
QY 71 DLPSDKRTTPEKIKONLAKPREQELKAVTENTSEKQITSGSQLESKESL-----SLN 125
Db 64 HLKSKR-----EERWKRKRQKKEKEMELHNIID-----DLLQKKSLCEVEELH 111
QY 126 KTVPTSNWEICDFTIKGNTLVGLSKSGVEK-----LSQT 160
Db 112 RTVQKQOQK--DFI--DGNVESLMTLEIEKSLKHEDIVDEIECIEKTLKRSSELRA 168
QY 161 DHLVPSQADGTQLIQVAFPTPK-KTAAIAYTS-----RAGENG 202
Db 169 DRLLAEAE-----SELSCTKEKTKNAVEKFTDAKRSLLQTESDAEELERRAQETA 218
QY 203 -----EISQDLDVGKEI-----INEGEV-----FNSYLLKKVTPTGYK 236
Db 219 VNLVKADQQLRSLQADAKDLQEHKIQEBILKEINKIVAAKSDDFQCLSKKKEKTEELQ 278
QY 237 HIGQDAFVNDKNIAYNLPESLETISDYAFALHAKQIDLPNLKA-----IGELAFPN 291
Db 279 KQKQDI-----EWAERNEHHLOVLKE-SEVLLQAKRAEL-EKLKQSVTSQQQWAVLDR 331
QY 292 Q-----ITGKLSLPRQLMLAERAP--KSNHKTIFRGNLSKLVIGEASF 334
Db 332 QLGHKKEELHLLQGSVVOAKADL-QEALRGTEVTEVCNKHIREVK-----SLLEELSF 384
QY 335 QNDNL-----SOLMLPDGLEKIESEAFNGDDHYNRVLVLTGSKGNPGLATEN 386
Db 385 QXGELNVQISERTQTLTK--QETBEKE-----ENLQVVLROMSKHTELK---- 429
QY 387 TVYNPDKLSQBSPEIDYTKWLEED---FTYOKNSVTGFSNK-----GLOKVRKNKLEIP 439
Db 430 ---NILDMLQENHELOGLK-LOHDORVSELEKTQVAVLEBKLELENLQISQOQKGELE 485
QY 440 KQHNGVTITEIGDANFRVNDPQNKTLR-----KYDLEEVKLPSTIRKIGA 484
Db 486 WQKQ---LLERDKREITERMTAESRALQSCVECLSEKEDLQEKCDIWEKKLAQTKRVLAA 542
QY 485 -----FAPQSNLKSFEAS-----DLLEEI-KEGAFMNNRIETLE-----LKDK 522
Db 543 AEENSKMEQSNLEKLELNVRLQELDLNRDKLSLHNDISAMQOQLOQEKREAVNSLOBE 602
QY 523 LVTIGDAAPHINHIYAIVLPESVOEIGRSAPQNGANNLI-FMGSVKVKTIGEMAFLSNLE 582
Db 603 LANQD---HLN-----LAKQDLHTTKHOD-----VLLSEQTRLOKQISEWANRFE 646
QY 583 HLDLSEQKQLTBPV--QAFSDNALKEVLLPASLKTIREEAFKQNLHLEKQLEVASALSHTA 640
Db 647 DCOKEEETKQOQLOVLQNEIEENGLKLVQEMMFQRLQKE--RESEESKLET----- 696
QY 641 FNALDNDGDEQFNKVVVK--THINSYALAGEHFIIVDPDKLSSTIVDLKILKI----- 695
Db 697 -----SKVTLKEQHOQLEKELTDQSKSL---DOVLSKVLAAEBERVRLTQBE 739
QY 696 -----EGLDYSLTROTTQTFQDMTTAGKALLSKSNLRQEGKQKFLQEA----- 739
Db 740 ERWCESLE-KTLSQTK-----RQUSEREQQLVEKS-----GELLALQKEADSMRADPSLRN 790
QY 740 QPFLGRVLDJKAIAKAEKALVTKKATPKNGQLLERSINKAVLAYNNSAIKK----- 789
Db 791 QFLTERKKAQKQVASKALKIORSQLEKNLLEQK-----QENSCIQKEMATIELVNAQ 843
QY 790 ---ANVRLEKELD 800
Db 844 DNHERRARLMKELN 857

RESULT 94

US-08-409-995-4

; Sequence 4, Application US/08409995

; Patent No. 5646259

GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen I.
APPLICANT: St. Geme III, Joseph W.
TITLE OF INVENTION: Haemophilus Adhesion Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RPT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

Query Match 3.1%; Score 158.5; DB 1; Length 1912;

Best Local Similarity 18.0%; Pred. No. 0.045;

Matches 211; Conservative 158; Mismatches 388; Indels 412; Gaps 55;

QY 47 YAESGSKLKINETSQPVDDTVTD-----LFSDKRTTPE-----KIKON 86
Db 566 HVEDAYKLLNLNKNANKQPLVTDSTAATVGDRLKLGWVSTKNGTKESNQVKQADE 625
QY 87 LAGPREBELKAVTEN-----TESEKQITSGSQLESKESLKNKTVPSTSWEICDFIT 141
Db 626 LFTGAGAATVTSKSENGKHTITVSVAETKADCGLEKGDGTIKLKVDNQNTDN-----VLT 680
QY 142 KGNTLVGLSKSGVE--KLSQTDH-----LVLPQOAA-----GTQLIQVASPAF 183
Db 681 VGNNGTAVTKGFEVTKGATDADRGRVTKDATTANDADKKVATVKDVATATNSAATFVK 740
QY 184 TPDKKTAAE-----YTSRAGENISQLDVGK-----EINIEGEVFN 222
Db 741 TENLFTSIDENPTDNGKDDALKAGDTLTFRAGKULKVKR---DGKNITFDLAKNLEVT 797
QY 223 SYLLKKVTI-----PTGYKHIGQDAFVDKNKTAENVLPESLETISDYAFALHAKQLDLPD 278
Db 798 AKVSDTLTIGGNTPTG---GTTA-----TPKVNITSTADGLN---FA-----KE 835
QY 279 NLKATGELAFPNQITGKLSLPRQLMLRAERAFKSNHIK----- 317
Db 836 TADASGSKNVYLGKATTLTLEP-----SAGAKSHVDNLVNDATKSKNAASIEDVLRA 888
QY 318 -TIEFRGNLSKVIG---EASFQDND-----LSQLMLPDG-----LEKIESA 355
Db 889 WNIQGNNGNVYVATYDVTNFTDDSTGTTTIVTQKADGKADVKIGAKTSVIKDHNGKL 948
QY 356 FTGNGPDHNNRVVLWTYKSKNP--SGLATENTYVNP--DKSLWQESPEIDYTKWLEEDFT 413
Db 949 FTGKDLKD--ANNGATVSEDDGDKDTGTGLVTAKTVIDAVNKGSGWRVTGE-----GA 997

```
QY 414 YQKNSVTGSKNGKQKVRKNKLEIPKQHNGVITTEIGDNAPRVDFQNKTLRYKDYLE-- 471
Db 998 TAETGATAVNAAGNAETVTSVGNF-KNGNATTATVSKDGNINV-----KYDVNVG 1048
QY 472 -----EVLKSTIRKIG-----AFAPQSNLKSFEASD 499
Db 1049 DGLKIGDDKKIVADTTTLTVTGKVSVPAGANSVNNKLVNAEGLATALLNL-SWTAKA 1107
QY 500 DL-----BEIKEG-----AFMNNRIE-----TLEKDKL-----VTIGDAAF 531
Db 1108 KYADGESEGETDQEVKAGDKVTFKAGKNLVKQSEKDFVSLQDTLTGLTSITLGGTAN 1167
QY 532 HINHIYAIVLPESVQ-EIGRSAPRONGANNLIPMGSKVKTIGEMAFLSNRLEHLDLSEQK 590
Db 1168 GRNDTGTVINKDGLTITILANGAAAGTDASN-----GNTISVTKOGISAGN-----K 1213
QY 591 QLTEIPVQAFSDNALKEVLLPASLKTIRREAPKKNHKLQLEVASALSH-----I 639
Db 1214 EITNV-----KSALTKYKDTQNTADETQDKFPAHVKNANEVEFVGKNGA 1258
QY 640 AFNALDDNDG-----DEQFDNKVVVKTHNSVALADGEHFIVDPDKLS 682
Db 1259 TVSAKTDNNKGHTVITIDVAEAKVGDGLEKDTGKIKLVDN-----TDGNNLLT----- 1307
QY 683 STIYDLEKILKLEGLDYSLRQTTQTOFRDMTTAGKALLSKN-----LRQK 730
Db 1308 ---VDATKGSVAGGEFNAVTTDTAGGTANERGVVVGSGNGATATETDKKKVATVG 1364
QY 731 EKQFLQBAQFQFLGRVDLDKAIAEAKALVTKKATKNGQLLERSINKAVLAYNNSAIKKA 790
Db 1365 DVAKAINDAATFV-KVEND-----DSATIDDSPTDDG-----ANDALKAXDTLTKAG 1411
QY 791 NVKALEKELDLTLGLVEGKGLAQMVMQGVYLLKPLPLPEYVIGLVNVPYFDSKGL-IY 849
Db 1412 ---KMLKVKRD-----GKNITPALANDLSVK 1434
QY 850 ALDMSDTIGEGQKDAYGNPILNVDEDNeg-----YHALAVA-TLADYEGLD 894
Db 1435 SATVSDKLSLGTN---GNKV-NITSDTKGLNPAKDSKGTGDDANIHLNGIASTLTD----- 1485
QY 895 IKTILNS-----KLSQTSIROVPTAAVHRAGIFQAIQNAIAAEAGLLPKP 940
Db 1486 --TLNLSGATNLGNGITDNEKKAASVKDVLNAGMNVGVKPPASANNQVENIDFVATY 1543
QY 941 GT-A--HSEKSSSESANSKDRGLQSNPK 965
Db 1544 DTVAFVSGDKDTSVTVESKDKGKRETVK 1572
```

RESULT 95

```
US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
```

```
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RPT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-685-467-4

Query Match 3.1%; Score 158.5; DB 3; Length 1912;
Best Local Similarity 18.0%; Pred. No. 0.045;
Matches 211; Conservative 158; Mismatches 388; Indels 412; Gaps 55;

QY 47 YAESGSKLKNITSGPVDVTD-----LFSDKRTTPE-----KIKDN 86
Db 566 HVEDAYGKLLNLEKNANKQLPLVTDSTAATVGDURKLGWVVSTKNGTKEESNOVKQADEV 625
QY 87 LAKPREQELKAVTEN-----TESEKQITSGSQLEQSKESLSLNTKTPSTSNWEICDPIT 141
Db 626 LFTGAGAATVTSKENGKHTITVSAETKADCGLEKQDFTIKLVKVDNQTUN-----VLT 680
QY 142 KGNTLVGLSGSVB-KLSQTDH-----LVLPSSAAD-----GTOLIQVASPAF 183
Db 681 VGNGTAVTKGPEVTKTGATDADRGKVTVKDATTANDADKKVATVKDVATAINSAATPVK 740
QY 184 TPDKKTAIAE-----YTSRAGENGESIQLDVVGK-----BIINEGEVFN 222
Db 741 TENLTTSIDEDNPTDNGKDDALKAGDTLTTFKAGKNLVKR---DGNKITFDLAKNLEVKT 797
QY 223 SYLLKVTI---PTGKHIGQDAFVDNKNIAEVLNPELFTISDYAFAPHALAQIDLPD 278
Db 798 AKVSDTUTIGNTPTG---GTTA-----TPKNITSTADGLN---FA-----KE 835
QY 279 NLKAIGELAPFNDQITGKLSLPQRLMRLAERAPKSNHIK-----TPKNITSTADGLN---FA----- 317
Db 836 TADASGSKNVYLKGIATTLTEP-----SAGAKSSHVDLNVDATKKSNAASIEDVLRA 888
QY 318 -TIEFRGNSLVIG---EASFQDND-----LSQMLPDPG-----LEKIBSEA 355
Db 889 WNIQNGNNDYVATYDTVNFDTDDSTGTTVTVTQKADGKADVKIGAKTSVIKDHNGKL 948
QY 356 FTGNPGDDHNNRVVLTAKGKNP-SGLATENTVNP-DKSLWQESPEIDYTKWLEEDFT 413
Db 949 FTGKDLKD-ANNGATVSEDDGKTGTGLVTKTAKTVIDAVNKGSMRVTEB-----GA 997
QY 414 YQKNSVTGSKNGKQKVRKNKLEIPKQHNGVITTEIGDNAPRVDFQNKTLRYKDYLE-- 471
Db 998 TAETGATAVNAAGNAETVTSVGNF-KNGNATTATVSKDGNINV-----KYDVNVG 1048
QY 472 -----EVLKSTIRKIG-----AFAPQSNLKSFEASD 499
Db 1049 DGLKIGDDKKIVADTTTLTVTGKVSVPAGANSVNNKLVNAEGLATALLNL-SWTAKA 1107
QY 500 DL-----BEIKEG-----AFMNNRIE-----TLEKDKL-----VTIGDAAF 531
Db 1108 KYADGESEGETDQEVKAGDKVTFKAGKNLVKQSEKDFVSLQDTLTGLTSITLGGTAN 1167
QY 532 HINHIYAIVLPESVQ-EIGRSAPRONGANNLIPMGSKVKTIGEMAFLSNRLEHLDLSEQK 590
Db 1168 GRNDTGTVINKDGLTITILANGAAAGTDASN-----GNTISVTKOGISAGN-----K 1213
```



```
QY 591 QLTETPVOAFSDNALKEVLLPASLKTIRREAFKKNHLKOLEVASALSH-----I 639
D 1214 EITNV-----KSAIKTYKDTQNTADETQKEFHAAVKNAVEFVKGNGA 1258
QY 640 AFNALDDNG-----DEQFQKVVVTHNSYALADGEHFIVDPDKLS 682
D 1259 TVSAKTDNNGKHTVTTIDVAEAKVGDGDKDTGDKIKLVKVDN-----TDGNLLT-----1307
QY 683 STIVLEKILKEGLDYSLRQTTQTPFRDMMTAGKALLSKN-----LRQ 730
D 1308 ---VDATKASVAKGEFNAVTTDATTAQGTANERKGVVKGSGNGATATETDKKVVATVG 1364
QY 731 EKQKFOEQAQFGRVLDLKAIAKAEKALVTKATKNGOLLERSINKAVLAYNNSAIKKA 790
D 1365 DVAKAINDAATFV-KVEND-----DSATIDDSPTDDG-----ANDALKAXDTLTKAG 1411
QY 791 NVKLEKEBLLDGLVEGKPLAQAATMVQGVLLKTPPLPEYIYGLNVYFKSGKL-IY 849
D 1412 --KLVKVRD-----GKNITFALANDLSVK 1434
QY 850 ALDMSDTIGEGOKDAYGNPILNVDENEG-----YHALAVA-TLADYEGLD 894
D 1435 SATVSDKLSLGTN---GNKV-NITSDTKGLNPAKDSKTGDDANIHLNGIASTLTD-----1485
QY 895 IKTIILNS-----KLSQTSIRQVPTAAHYHRAGIFQAIQNAAAAEQQLPKP 940
D 1486 --TLNLSGATTNLGNGITDNEKKAASVKDVLNAGMNVGVKVPASANNQVENIDFVATY 1543
QY 941 GT-----HSEKSSSESANSKORGLQSNPK 965
D 1544 DTVDVFGDKDTTSVTVESKONGKRTVEK 1572

RESULT 96
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-33

Query Match 3.1%; Score 158.5; DB 3; Length 2353;
Best Local Similarity 18.2%; Pred. No. 0.062;
Matches 213; Conservative 155; Mismatches 389; Indels 412; Gaps 55;

QY 47 YABSSGSKLKINETSGPDDTVD-----LFSKRTTPE-----KIKON 86
D 567 HVEDAYKGLNLNNEKANKQPLVTDSTAATVGDRLKLGHVSTKNGTKEESQVQKQADEV 626
QY 87 LAGPREQELKAVTEN-----TESEKQITSGSQLESLSLNTKVPSTSNWEICDFIT 141
D 627 LFTGAGAAVTSKSNGKHTITVSAETKADCGLEKDGDTIKLVKDNQNTDN-----VLT 681
QY 142 KGNLTVGLSKSGVE--KLSQTDH-----LVLSQAAD-----GTQLIQVASFAF 183
D 682 VGNNGPAVTGKFVTKATDADRGKVTVKQATANDADKKVATVKQDVATNAISAATFVK 741
```

```
QY 184 TPDKKTAAE-----YTSRAGEGEISQLDVGK-----EINIAGEVFN 222
D 742 TENLTTSIDEDNPTDNGKDDALKAGDTLLTFKAGKNLKVKR-----DGKNIITFDLAKNLKVT 798
QY 223 SYLLKKVIT-----PTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFAPHAHLKALQIDLDP 278
D 799 AKVSDTLTIGGNTPTG---GTTA-----TPKVNITSTADGLN---FA-----KE 836
QY 279 NLKAIGELAFPDNQITGKLSLPRQLMRLAERAFKSNHIK-----317
D 837 TADASGSKNVYKGLATLTPE-----SAGAKSHVDLNVDAATKKSNAASIEDVLRAG 889
QY 318 -TIFRGNLSKVIG---EASFQDND-----LSQLMLPDG-----LEKTESRA 355
D 890 WNIQNGNNVDVATVDTVNFDTDDSTGTTVTQKADGKADGVKIGAKTSVTKDHNGKL 949
QY 356 FTGNGDGDHYNRVVLWTKSGKNP-SGLATENTYNP-DKSLWQSPEDITYFKWLEEDFT 413
D 950 FTGKDLKD-ANNGATVSEDDGKDTGTGLVTAKTVJDAVNKSGMRVTGE-----GA 998
QY 414 YOKNSVTGFSNKGLOKVKRKNKLEIPKQHNGVTITEIGDQNAFRNVDFQNKTLRKVDLE-- 471
D 999 TAETGATAVNAGNAETVTSGETSVNF-KNGNATTATVSKDNGNINV-----KVDNVVG 1049
QY 472 -----EVLKPFSTIRKIG-----AFAFQSNLKSFEASD 499
D 1050 DGLKIGDDKKIVADTTTLTVTGKVSVPAGANSVNNKKLVNAEGLATALNLL-SWTAKA 1108
QY 500 DL-----EIKEG-----AFMNRIE-----TLBLKOKL-----VTTGDAAP 531
D 1109 DKYADGESEGETDQSVKAGDKVTFKAGKNLKVQSEKDFYSLQDTLTGLTSITLGGTAN 1168
QY 532 HINHIYAVLVPESVQ-EIGRSAFRQNGANNLIFMGSKVTKLGEMAFSLNHLHLDSEOK 590
D 1169 GRNDGTGVNKDGLTITLANGAAAGTDASN-----GNTISVTKDGISAGN-----K 1214
QY 591 QLTETPVOAFSDNALKEVLLPASLKTIRREAFKKNHLKOLEVASALSH-----I 639
D 1215 EITNV-----KSAIKTYKDTQNTADETQKEFHAAVKNAVEFVKGNGA 1259
QY 640 AFNALDDNG-----DEQFQKVVVTHNSYALADGEHFIVDPDKLS 682
D 1260 TVSAKTDNNGKHTVTTIDVAEAKVGDGDKDTGDKIKLVKVDN-----TDGNLLT-----1308
QY 683 STIVLEKILKEGLDYSLRQTTQTPFRDMMTAGKALLSKN-----LRQ 730
D 1309 ---VDATKASVAKGEFNAVTTDATTAQGTANERKGVVKGSGNGATATETDKKVVATVG 1365
QY 731 EKQKFOEQAQFGRVLDLKAIAKAEKALVTKATKNGOLLERSINKAVLAYNNSAIKKA 790
D 1366 DVAKAINDAATFV-KVEND-----DSATIDDSPTDDG-----NDALKAG 1404
QY 791 NVKLEKEBLLDGLVEGKPLAQAATMVQGVLLKTPPLPEYIYGLNVYFKSGKL-IY 849
D 1405 -----DTLT-----LKAGKNLKVKRDGKNITFALANDLSVK 1435
QY 850 ALDMSDTIGEGOKDAYGNPILNVDENEG-----YHALAVA-TLADYEGLD 894
D 1436 SATVSDKLSLGTN---GNKV-NITSDTKGLNPAKDSKTGDDANIHLNGIASTLTD-----1486
QY 895 IKTIILNS-----KLSQTSIRQVPTAAHYHRAGIFQAIQNAAAAEQQLPKP 940
D 1487 --TLNLSGATTNLGNGITDNEKKAASVKDVLNAGMNVGVKVPASANNQVENIDFVATY 1544
QY 941 GT-----HSEKSSSESANSKORGLQSNPK 965
D 1545 DTVDVFGDKDTTSVTVESKONGKRTVEK 1573
```

RESULT 97
US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578

[illegible]

```
Db      890 WNIQGNVNDVAVYDVTNFTDDSTGTTTIVTQKADGKADVKIGAKTSVVKDHNGKL 949
Qy      356 FTGNGPDHNNRVVLTWTKSGKNP-SGLATENTYVNP-DKSLWQBSPEIDYTKWLEEDFT 413
Db      950 FTGDKLKD-ANNGATVSEDDGKTGTGLVTAKTVIDAVNKGSRVTGE-----GA 998
Qy      414 YQKNSVTGFSNKGLOKVRKNKLEIPKQHNGVTITEIGDNAPRVDFQNKTLRKVDLE-- 471
Db      999 TAETGATAVNAGNAETVTSVNF-KNGNATTATVSKDNGNINV-----KYDVNVG 1049
Qy      472 -----EVKLPSTIRKIG-----AFAFQSNLNKSFASD 499
Db      1050 DGLKIGDDKKIVADTTTLTVTGKVSVPAGANSVNNKKLVNAEGLATLNNL-SWTAKA 1108
Qy      500 DL-----EBIKEG-----AFMNRIE-----TLELKKL-----VTIGDAAF 531
Db      1109 DKYADGSEGETDQSVKAGDKVTTFKAGKNLVKQSEKDFTSYSLQDTLGLTSITLGGTAN 1168
Qy      532 HINHYIAVLPSVQ-EIGRSAPFRONGANNLIFPMGSKVKTIGEMAFPLNRLHLDLSEOK 590
Db      1169 GRNDTGTWINKDGLTITLANGAAAGTASN-----GNTISVTKDGISAGN-----K 1214
Qy      591 OLTEIPVQAFSDNALKEVLLPASLTIREAFKKNHLKQLEVASALSH-----I 639
Db      1215 EITNV-----KSALETYKDTQNTADETQKEFHAAVKNAVEFVKGKGA 1259
Qy      640 AFNALDDNDG-----DEQFNKVVVTKHNSYALADGEHPIDVPDKLS 682
Db      1260 TVSAKTNNNGKHTVIDVAEAKVGDGLEDKTDGKIKLVKN-----TDGNLLT----- 1308
Qy      683 STIVOLEKILKLEGLDVTSLRQTTQTPRDMTTAGKALLSKN-----LRQK 730
Db      1309 ---VDATGASVAKGEFAVTTTATTAQGTNANERGVVVGSGNGATATETDKKVVATVG 1365
Qy      731 EKQFLQBAQFPLGRVDLDKALAKAEKALVTKKATKNGQLLERSINKAVLAVNSAIKKA 790
Db      1366 DVAKAINDAATFV-KVEND-----DSATIDDSPTDDGA-----NDALKAG 1404
Qy      791 NVKLEKELDLTLGLVEGKGPLAQATWQGVVLLKPLPLPEYITGLNVYFDKSGKL-IY 849
Db      1405 -----DTLT-----LKAGKNLVKVRDGNKITFALANDLSVK 1435
Qy      850 ALDMSDTTGEQKDAYGNPILNVDENEG-----YHALAVA-TLADYEGLD 894
Db      1436 SATVSDKLSLGTN-----GNKV-NITSDTKGLNPAKDSKTGDDDANIHLNGIASTLTD----- 1486
Qy      895 IKTILNS-----KLSQLTSIRQVPTAAVHRAGIFQAIQNAAEAEQLPKP 940
Db      1487 ---TLNNGATNVLGNGITDNEKKRAASVKVNLNAGNVRGVKPSANNOVENIDFVATY 1544
Qy      941 GT-----HSEKSSSESANSKORGLQSNPK 965
Db      1545 DTVPVSGDKOTTSVTVESKONGKTEVK 1573
```

RESULT 100

```
US-09-684-707-4
; Sequence 4, Application US/09684707
; Patent No. 6759213
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph
; Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; City: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/684,707
; FILING DATE: 04-Oct-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-684-707-4
```

```
Query Match      3.1%; Score 158.5; DB 4; Length 2353;
Best Local Similarity 18.2%; Pred. No. 0.062;
Matches 213; Conservative 155; Mismatches 389; Indels 412; Gaps 55;

Qy      47 YAESGSGSKLKINETSQVDDTVTD-----LPSDKRTTPE-----KIKDN 86
Db      567 HVEDAYKGLMLNENKANKQPLVTDSTAATVGDRLKLGWVVSTKNGTKESNQVKADEV 626
Qy      87 LAKGPREQELKAVTEN-----TESEKQITSGSQLEQSKESLSLNKTVSPSTNWEICDPT 141
Db      627 LPTGAGAATVTSKSENGKHITTVSAETKADCGLEKDGDTIKLVQDNQNTDN-----VLT 681
Qy      142 KGNTLVLSKSGVE--KLSQTDH-----LVLPQAAAD-----GTOLIQVASPAF 183
Db      682 VGNNGTAVTKGPFETVKGTADADRGKVTVDATANDADKKVATVKDVATAINSAAATPVK 741
Qy      184 TPDKKTAAIE-----YTSBAGENGESISQLDVGK-----EIIIEGEVFN 222
Db      742 TENLTTSIDEDNPTDNGKODALKAGDTLTFKAGKNLVKVR---DGKNITFDLAKNLEVK 798
Qy      223 SYLLKVKVTI---PTGVKHIGQDAFVDNKNIAEYVNLPESETISDYAFALHAKQIDLPD 278
Db      799 AKVSDTLTIGGNTPTG---GTTA-----TPKNITSTADGLN---FA-----KE 836
Qy      279 NLKAIGELAPFDNQITGKLSLPROLMRLAERAFKSNHIK-----317
Db      837 TADAGSGKNVYLKGIATTLTPE-----SAGAKSSHVDLNVNDAATKKSNAAASIEDVLRAG 889
Qy      318 -TIEFRGNLSKVIG---EASFQDND-----LSQLMLPDG-----LEKISEA 355
Db      890 WNIQGNVNDVAVYDVTNFTDDSTGTTTIVTQKADGKADVKIGAKTSVVKDHNGKL 949
Qy      356 FTGNGPDHNNRVVLTWTKSGKNP-SGLATENTYVNP-DKSLWQBSPEIDYTKWLEEDFT 413
Db      950 FTGDKLKD-ANNGATVSEDDGKTGTGLVTAKTVIDAVNKGSRVTGE-----GA 998
Qy      414 YQKNSVTGFSNKGLOKVRKNKLEIPKQHNGVTITEIGDNAPRVDFQNKTLRKVDLE-- 471
Db      999 TAETGATAVNAGNAETVTSVNF-KNGNATTATVSKDNGNINV-----KYDVNVG 1049
Qy      472 -----EVKLPSTIRKIG-----AFAFQSNLNKSFASD 499
```

Db	1050	DGLKIGDDKKIADTTTLT	TG	GK	V	P	G	A	N	N	K	L	V	N	E	G	L	A	T	A	L	N	N	L	-	S	M	T	A	K	1108																								
Qy	500	DL-----	BEI	KEG	-----	AF	M	N	R	I	E	-----	T	L	E	K	D	K	L	-----	V	T	I	G	D	A	A	F	531																										
Db	1109	DKYADGESEGTQ	DE	V	R	K	A	G	D	K	V	T	F	K	A	G	N	L	K	V	K	Q	S	E	K	D	F	T	S	L	Q	D	T	L	T	G	L	T	S	I	T	L	G	T	A	N	1168								
Qy	532	HINHIYAIVLP	ES	V	Q	-	E	I	R	S	A	P	Q	N	C	A	N	N	L	I	P	M	G	S	K	V	T	L	G	E	M	A	F	I	N	S	E	L	H	L	D	I	S	E	Q	590									
Db	1169	GRNDTGIV	I	N	K	G	T	I	I	L	A	N	A	A	G	T	A	S	N	-----	G	N	T	I	S	T	V	K	O	I	S	A	G	N	-----	K	1214																		
Qy	591	QLTEIPVQAF	S	D	N	A	L	K	E	V	L	P	A	S	I	K	T	I	R	E	A	F	K	N	H	L	K	O	L	E	V	A	S	A	L	S	H	-----	I	639															
Db	1215	EITNV-----	KS	A	L	T	K	Y	D	T	O	N	T	A	E	T	D	Q	K	E	F	A	A	V	K	N	A	N	E	V	E	F	V	G	K	N	A	1259																	
Qy	640	AFNALDDNG	-----	DE	Q	F	D	N	K	V	V	V	T	H	N	S	A	L	A	D	G	E	H	F	I	V	D	P	D	K	L	S	682																						
Db	1260	TVS	A	K	T	D	N	N	G	H	V	T	I	D	V	A	E	A	K	V	G	D	E	K	O	D	G	K	I	K	L	K	V	D	N	-----	T	D	G	N	N	L	L	T	-----	1308									
Qy	683	STIVDLBKIL	I	B	G	L	D	S	T	L	R	O	T	O	T	Q	F	R	D	M	T	T	A	G	R	A	L	L	S	K	N	-----	L	R	O	Q	730																		
Db	1309	---	V	D	A	T	G	S	A	V	A	K	E	F	N	A	V	T	T	D	A	T	A	Q	G	T	N	A	N	E	R	G	K	V	V	V	K	S	N	G	A	T	A	T	E	T	D	K	K	V	A	T	V	G	1365
Qy	731	EKQFLQEAQ	O	F	L	G	R	V	D	L	D	K	A	I	A	E	A	L	V	T	K	A	T	K	N	G	O	L	L	E	R	S	I	N	K	A	V	L	A	N	S	A	I	K	A	790									
Db	1366	DVAKAINDA	T	F	V	-	K	V	E	N	-----	DS	A	T	I	D	D	S	E	T	D	D	G	A	-----	N	D	K	A	K	A	1404																							
Qy	791	NVRLEKE	L	D	L	L	T	G	L	V	E	G	K	G	P	L	A	Q	A	T	M	V	O	G	V	Y	L	L	K	T	P	L	P	E	V	Y	I	G	L	N	V	F	D	S	G	K	L	-	I	Y	849				
Db	1405	-----	D	T	L	T	-----	L	K	A	G	N	L	K	V	K	R	D	G	K	N	I	T	F	A	L	N	D	L	S	V	K	1435																						
Qy	850	ALOWSDT	I	G	E	G	O	K	D	A	Y	G	N	P	I	N	V	E	D	E	G	-----	Y	H	A	L	A	V	A	-	T	L	A	D	E	G	L	D	894																
Db	1436	SATVSD	K	L	S	L	G	T	N	-----	G	N	K	V	-	N	I	T	S	D	T	K	G	L	N	F	A	R	K	S	T	G	D	A	N	I	H	L	N	G	I	A	S	T	L	T	-----	1486							
Qy	895	I	K	T	I	N	S	-----	K	L	S	O	L	T	S	I	R	O	V	P	T	A	A	V	H	R	A	G	I	F	O	A	T	O	N	A	A	A																	

Search completed: August 28, 2005, 10:50:01
Job time : 88 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:50:07 ; Search time 182 Seconds
(without alignments)
2836.131 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 1008

Sequence: 1 MKHLKLTVALTTTTSVWTH.....LGYTSVALLTAIKKKY 1008

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 110 summaries

Database : Uniprot_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	100.0	1008	2 Q9A0C0	Q9A0C0 streptococ
2	679	67.4	1008	2 Q8P1F7	Q8P1F7 streptococ
3	293	29.1	1008	2 Q878I7	Q878I7 streptococ
4	283	28.1	999	2 Q8K7X8	Q8K7X8 streptococ
5	89	8.8	1055	2 Q8S6V3	Q8S6V3 streptococ
6	76	7.5	1055	2 Q8ELD6	Q8ELD6 streptococ
7	9	0.9	245	2 Q9ZRT5	Q9ZRT5 arabidopsis
8	9	0.9	279	2 Q9NW78	Q9NW78 homo sapien
9	9	0.9	455	2 Q86A18	Q86A18 dictyosteli
10	9	0.9	523	2 Q8CL34	Q8CL34 mus musculu
11	9	0.9	564	2 Q80TA6	Q80TA6 mus musculu
12	9	0.9	568	2 Q9D2Y7	Q9D2Y7 mus musculu
13	9	0.9	583	2 Q86X27	Q86X27 homo sapien
14	9	0.9	590	2 Q9FHE1	Q9FHE1 arabidopsis
15	9	0.9	590	2 Q7TPZ6	Q7TPZ6 mus musculu
16	9	0.9	590	2 Q8SZU2	Q8SZU2 mus musculu
17	9	0.9	590	2 Q9ERD6	Q9ERD6 mus musculu
18	9	0.9	907	1 LGR5_HUMAN	Q75473 homo sapien
19	8	0.8	12	2 Q75MW9	Q75MW9 homo sapien
20	8	0.8	67	2 Q982N8	Q982N8 rhizobium
21	8	0.8	98	2 Q7UVJ9	Q7UVJ9 rhodospirell
22	8	0.8	106	1 RL3E_SULTO	P58376 sulfobolus
23	8	0.8	116	2 Q9NSZ8	Q9NSZ8 homo sapien
24	8	0.8	165	2 Q87WF0	Q87WF0 pseudomonas
25	8	0.8	174	2 Q9SB23	Q9SB23 hordeum vul
26	8	0.8	237	2 Q9FR02	Q9FR02 nicotiana t
27	8	0.8	240	2 Q8TW63	Q8TW63 methanopyru
28	8	0.8	241	2 Q7T3A2	Q7T3A2 brachydanio
29	8	0.8	242	2 Q7NST9	Q7NST9 chromobacte
30	8	0.8	253	2 P93664	P93664 chlamydomon
31	8	0.8	255	2 Q8AB78	Q8AB78 bacteroides

32	8	0.8	255	2 Q8UYE5	Q8UYE5 camelpox vi
33	8	0.8	255	2 Q6YLZ5	Q6YLZ5 camelpox vi
34	8	0.8	255	2 Q776C0	Q776C0 camelpox vi
35	8	0.8	268	2 Q9H0J8	Q9H0J8 homo sapien
36	8	0.8	284	2 Q96191	Q96191 plasmodium
37	8	0.8	296	2 Q926G6	Q926G6 rhizobium m
38	8	0.8	304	2 Q9ABZ6	Q9ABZ6 caulobacter
39	8	0.8	316	2 Q88U35	Q88U35 lactobacill
40	8	0.8	329	1 YDZA_SCHPO	O13717 schizosacch
41	8	0.8	348	2 Q99Z7	Q99Z7 cyprinella
42	8	0.8	348	2 Q99944	Q99944 erimonax mo
43	8	0.8	348	2 Q8M518	Q8M518 pelvichacho
44	8	0.8	348	2 Q8HLA4	Q8HLA4 indostomus
45	8	0.8	351	2 Q26490	Q26490 spodoptera
46	8	0.8	360	2 Q8RPN6	Q8RPN6 legionella
47	8	0.8	360	2 Q9F7S2	Q9F7S2 gamma-prote
48	8	0.8	370	2 Q87S80	Q87S80 vibrio para
49	8	0.8	385	2 Q7QF26	Q7QF26 anopheles g
50	8	0.8	391	2 Q63C90	Q63C90 bacillus ce
51	8	0.8	391	2 Q739H8	Q739H8 bacillus ce
52	8	0.8	391	2 Q81RG8	Q81RG8 bacillus an
53	8	0.8	391	2 Q6HJ01	Q6HJ01 bacillus th
54	8	0.8	393	1 ACK2_RHIME	Q9X449 rhizobium m
55	8	0.8	416	2 Q6BJ95	Q6BJ95 debaryomyce
56	8	0.8	427	1 AROA_STRMU	Q8DUV8 streptococ
57	8	0.8	427	2 Q8CCK5	Q8CCK5 mus musculu
58	8	0.8	429	2 Q97KA1	Q97KA1 clostridium
59	8	0.8	441	2 Q9NKL1	Q9NKL1 dictyosteli
60	8	0.8	458	1 ARDY_BACHD	Q9X821 bacillus ha
61	8	0.8	463	2 Q7VVB6	Q7VVB6 bordetella
62	8	0.8	463	2 Q7WKU2	Q7WKU2 bordetella
63	8	0.8	463	2 Q7W7F4	Q7W7F4 bordetella
64	8	0.8	478	2 Q9VE79	Q9VE79 aeropyrum p
65	8	0.8	483	2 Q9NW94	Q9NW94 homo sapien
66	8	0.8	485	2 Q6LUN0	Q6LUN0 photobacter
67	8	0.8	488	2 Q42989	Q42989 oryza sativ
68	8	0.8	488	2 Q42990	Q42990 oryza sativ
69	8	0.8	488	2 Q6Z5B2	Q6Z5B2 oryza sativ
70	8	0.8	490	2 Q8WPP2	Q8WPP2 trypanoplas
71	8	0.8	492	2 Q6C9U7	Q6C9U7 yarrowia li
72	8	0.8	503	1 AMYB_WHEAT	P93594 triticum ae
73	8	0.8	503	1 GPMY_MYCGA	Q7NAQ5 mycoplasma
74	8	0.8	503	2 Q88335	Q88335 secale cere
75	8	0.8	517	2 Q6SNP7	Q6SNP7 hordeum vul
76	8	0.8	528	2 Q7MD66	Q7MD66 vibrio vuln
77	8	0.8	533	2 Q9FUK6	Q9FUK6 hordeum vul
78	8	0.8	533	2 Q9FUK7	Q9FUK7 hordeum vul
79	8	0.8	535	1 AMYB_HORVU	P16098 hordeum vul
80	8	0.8	535	2 P82993	P82993 hordeum spo
81	8	0.8	535	2 Q84T19	Q84T19 hordeum vul
82	8	0.8	535	2 Q84T20	Q84T20 hordeum vul
83	8	0.8	535	2 Q9AVJ8	Q9AVJ8 hordeum vul
84	8	0.8	535	2 Q9FSI3	Q9FSI3 hordeum vul
85	8	0.8	535	2 Q9SBH7	Q9SBH7 hordeum vul
86	8	0.8	543	2 Q7X3X6	Q7X3X6 clavibacter
87	8	0.8	546	2 Q6VLX2	Q6VLX2 homo sapien
88	8	0.8	554	2 Q8WKW9	Q8WKW9 xenopus lae
89	8	0.8	555	2 Q38914	Q38914 arabidopsis
90	8	0.8	555	2 Q42462	Q42462 arabidopsis
91	8	0.8	594	2 Q9P4A3	Q9P4A3 emericella
92	8	0.8	625	2 Q7VQZ4	Q7VQZ4 candidatus
93	8	0.8	642	2 Q9FFV8	Q9FFV8 arabidopsis
94	8	0.8	649	2 Q6ZPU7	Q6ZPU7 mus musculu
95	8	0.8	662	2 Q6MW74	Q6MW74 oryza sativ
96	8	0.8	673	2 Q7XP38	Q7XP38 oryza sativ
97	8	0.8	681	2 Q7Z5B6	Q7Z5B6 homo sapien
98	8	0.8	681	2 Q54781	Q54781 mus musculu
99	8	0.8	682	2 Q8VCD9	Q8VCD9 mus musculu
100	8	0.8	686	2 P78362	P78362 homo sapien
101	8	0.8	688	2 Q81YQ3	Q81YQ3 homo sapien
102	8	0.8	688	2 Q6NULO	Q6NULO homo sapien
103	8	0.8	719	2 Q7NRC2	Q7NRC2 chromobacte
104	8	0.8	757	2 Q98PD0	Q98PD0 rhizobium l

105 8 0.8 776 2 Q9AAE9 Q9aae9 caulobacter
 106 8 0.8 784 1 TL32 MACFA Q95m53 macaca fasc
 107 8 0.8 819 2 Q8CIG4 Q8cig4 mus musculus
 108 8 0.8 862 2 Q9P5X6 Q9p5x6 neurospora
 109 8 0.8 879 2 Q76978 Q76978 scypha raph
 110 8 0.8 880 1 SYV_BACST P11931 bacillus st

RESULT 1
 Q9AAC0
 ID Q9AAC0 PRELIMINARY; PRT; 1008 AA.
 AC Q9AAC0;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein SPY0843.
 GN OrderedLocusNames=SPY0843;
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071593998;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
 DR EMBL; AE006534; AAK33772.1; -;
 DR GO; GO:000986; C:cell surface; IEA.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR007093; LRR_Tp.
 DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 1008 AA; 111503 MW; 0458P8CE931AF0CF CRC64;

Query Match 100.0%; Score 1008; DB 2; Length 1008;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKLKLTVALTITTVSVTHNQEVSLVKEPIKQTSASSISGADYAESGSKLKINE 60
 Db 1 MKKLKLTVALTITTVSVTHNQEVSLVKEPIKQTSASSISGADYAESGSKLKINE 60

QY 61 TSGPVDVDTDLFSDKRTPEIKDKNLAKPREQELKAVTENSEKQITSGSQLEQSK 120
 Db 61 TSGPVDVDTDLFSDKRTPEIKDKNLAKPREQELKAVTENSEKQITSGSQLEQSK 120

QY 121 SLSLNKTPVSTNWEICDITKGNLTGLVLSKGVKLSQTDHLVLPQAADGTQLIQVAS 180
 Db 121 SLSLNKTPVSTNWEICDITKGNLTGLVLSKGVKLSQTDHLVLPQAADGTQLIQVAS 180

QY 181 FAFTPDKKTAAIAYTSRAGENGESIQDVGKEIINEGEVFNYSLLKKVTPTGYKHGQ 240
 Db 181 FAFTPDKKTAAIAYTSRAGENGESIQDVGKEIINEGEVFNYSLLKKVTPTGYKHGQ 240

QY 241 DAFVQNKIAEVLNPLESLETISDYAPAHALKQIDLPNLKAIGBLAFPDNQITKLSLP 300
 Db 241 DAFVQNKIAEVLNPLESLETISDYAPAHALKQIDLPNLKAIGBLAFPDNQITKLSLP 300

QY 301 ROLMLAERAFKSNHKTTEFGNSIKVIGEASFOQNDLSQMLPDGLEKTESAFTGNP 360
 Db 301 ROLMLAERAFKSNHKTTEFGNSIKVIGEASFOQNDLSQMLPDGLEKTESAFTGNP 360

QY 361 GDDHYNNRVLVTKGKQPSGLATENTYVNPDKSLWQESPEIDYTKLBEEDFTYQKNSVT 420
 Db 361 GDDHYNNRVLVTKGKQPSGLATENTYVNPDKSLWQESPEIDYTKLBEEDFTYQKNSVT 420

QY 421 GFNSKGLQKVRKNLEIPKQHNQVTTIEIGDNAFRNVDPQNKTLRKVDLEEVKLPSTIR 480
 Db 421 GFNSKGLQKVRKNLEIPKQHNQVTTIEIGDNAFRNVDPQNKTLRKVDLEEVKLPSTIR 480

QY 481 KIGAFAPQSNLKGFEASDDLEIEKEGAFNMNRITETLEKDKLVITIGDAAAFHINHIYAIV 540
 Db 481 KIGAFAPQSNLKGFEASDDLEIEKEGAFNMNRITETLEKDKLVITIGDAAAFHINHIYAIV 540

QY 541 LPESVQIEGRSAFRQNGANNLIIFMGSKVKTIGEMAFLSNRLEHLDLSEKQLTPTIPVQAF 600
 Db 541 LPESVQIEGRSAFRQNGANNLIIFMGSKVKTIGEMAFLSNRLEHLDLSEKQLTPTIPVQAF 600

QY 601 SDNALKEVLLPASLKTITREAFKGNHLKQLEVASALSHIAFNALDDNDGDQFQNKVVVK 660
 Db 601 SDNALKEVLLPASLKTITREAFKGNHLKQLEVASALSHIAFNALDDNDGDQFQNKVVVK 660

QY 661 THNSYALADGEHFIVDPDKLSSTIVDEKILKIEGLDYSTLQTTQTTQTFQFRDMTTAGKA 720
 Db 661 THNSYALADGEHFIVDPDKLSSTIVDEKILKIEGLDYSTLQTTQTTQTFQFRDMTTAGKA 720

QY 721 LLSKSNLRQGEKQFLQEAQFPLGRVLDLKAIAKAERKALVTKATQKQLLERSINKAVL 780
 Db 721 LLSKSNLRQGEKQFLQEAQFPLGRVLDLKAIAKAERKALVTKATQKQLLERSINKAVL 780

QY 781 AYNSAIKKNVKNLEKELDLTLTGLVSGKGPLAQATWQGVYLLKTPPLPEYYIGLVNY 840
 Db 781 AYNSAIKKNVKNLEKELDLTLTGLVSGKGPLAQATWQGVYLLKTPPLPEYYIGLVNY 840

QY 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDSNDEGYHALAVATLADYEGLDIKTLN 900
 Db 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDSNDEGYHALAVATLADYEGLDIKTLN 900

QY 901 SKLSQLTSSIRQVPTAAVHRAGIFQAIQNAAEAEQLLPKGTSEKSSSSSSANSKDRGL 960
 Db 901 SKLSQLTSSIRQVPTAAVHRAGIFQAIQNAAEAEQLLPKGTSEKSSSSSSANSKDRGL 960

QY 961 QSNPKTNRGRHSAILPRTGSGKSFYVILGTVTSVALLSLITAIKKKKY 1008
 Db 961 QSNPKTNRGRHSAILPRTGSGKSFYVILGTVTSVALLSLITAIKKKKY 1008

RESULT 2
 Q8P1F7
 ID Q8P1F7 PRELIMINARY; PRT; 1008 AA.
 AC Q8P1F7;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein spyM18_0903.
 GN OrderedLocusNames=spyM18_0903;
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232;
 RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy L.G., Mueser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 group A Streptococcus strains associated with acute rheumatic fever
 outbreaks";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
 DR EMBL; AE010019; AAL9753.1; -;
 DR GO; GO:000986; C:cell surface; IEA.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR007093; LRR_Tp.
 DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
 KW Complete proteome; Hypothetical protein.

SQL	SEQUENCE	i008 AA; 111620 MW; 89E79683FC2F03B CRC64;
	Query Match	67.4%; Score 679; DB 2; Length 1008;
	Best Local Similarity	99.8%; Pred. No. 0;
	Matches	879; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	128	VPSTSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHVLVPSQAADGTQLIQVASFAPTPDK 187
DB	128	VPSTSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHVLVPSQAADGTQLIQVASFAPTPDK 187
QY	188	KTAIABYTSRAGENGISOLDVDGKEIINEGEVFNYSLLKKVTIPYGVKHIGODAFVNDK 247
DB	188	KTAIABYTSRAGENGISOLDVDGKEIINEGEVFNYSLLKKVTIPYGVKHIGODAFVNDK 247
QY	248	NIAEVLNPSLETISDYAFALHAKQIDLPDNLKAIGELAFDNDQITGKLSLPRQLMRLA 307
DB	248	NIAEVLNPSLETISDYAFALHAKQIDLPDNLKAIGELAFDNDQITGKLSLPRQLMRLA 307
QY	308	ERAPKSNHIKTIEFRGNLSKVIGEASFQNDLSQLMPLDGLKIESEAPTGNPGDDHNN 367
DB	308	ERAPKSNHIKTIEFRGNLSKVIGEASFQNDLSQLMPLDGLKIESEAPTGNPGDDHNN 367
QY	368	RVLVMTKSKNPGSLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYQKNSVTGSSKGL 427
DB	368	RVLVMTKSKNPGSLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYQKNSVTGSSKGL 427
QY	428	QVKRKNLKEIPKQHNGVTITEIGDGNFRNVDFQNTLRKYDLEEVKLPSTIRKIGAPAF 487
DB	428	QVKRKNLKEIPKQHNGVTITEIGDGNFRNVDFQNTLRKYDLEEVKLPSTIRKIGAPAF 487
QY	488	QSNLKSFPASDDLEEKEGAFMNNRIETLEKDKLVITIGDAAFINHIYAIVLPSVQ 547
DB	488	QSNLKSFPASDDLEEKEGAFMNNRIETLEKDKLVITIGDAAFINHIYAIVLPSVQ 547
QY	548	IGRSAPFRONGANNLIPMGSKVKTIGEMAFISNRLEHLDLSEQKLTETIPVQAFSDNALKE 607
DB	548	IGRSAPFRONGANNLIPMGSKVKTIGEMAFISNRLEHLDLSEQKLTETIPVQAFSDNALKE 607
QY	608	VLLPASLKTIRBEAFKKNHLKQEVASALSHAFNALDNDGDEQFNDKVVVKTTHNSYA 667
DB	608	VLLPASLKTIRBEAFKKNHLKQEVASALSHAFNALDNDGDEQFNDKVVVKTTHNSYA 667
QY	668	LADGEHFI VDPDKLSTI VDLBKILKLEGLDYSTLRQTQTQPRDMTTAGKALLSKSNL 727
DB	668	LADGEHFI VDPDKLSTI VDLBKILKLEGLDYSTLRQTQTQPRDMTTAGKALLSKSNL 727
QY	728	RODEKQKFLQEAQFPLGRVLDLKA TAKAEKALVTKATKNGQLLERSINKAVLAYNNSAI 787
DB	728	RODEKQKFLQEAQFPLGRVLDLKA TAKAEKALVTKATKNGQLLERSINKAVLAYNNSAI 787
QY	788	KKANVKRLKELDLTLTGLVEGKPLAQATMVQGVYLLKTPPLPEYIIGLVVYFDKSGKL 847
DB	788	KKANVKRLKELDLTLTGLVEGKPLAQATMVQGVYLLKTPPLPEYIIGLVVYFDKSGKL 847
QY	848	IYALDMSDTIGECQKDAYGNPILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQLT 907
DB	848	IYALDMSDTIGECQKDAYGNPILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQLT 907
QY	908	SIRQVPTAAYHRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGLQSNPKTN 967
DB	908	SIRQVPTAAYHRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGLQSNPKTN 967
QY	968	RGHSAILLPRTGSKGSFVYGIIGYTSVALLSLITAIKKKY 1008
DB	968	RGHSAILLPRTGSKGSFVYGIIGYTSVALLSLITAIKKKY 1008

RESULT 3

ID Q87817

AC Q87817; PRELIMINARY; PRT; 1008 AA.

DT 01-JUN-2003 (TremBLrel. 24, Created)

DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)

DT	01-MAR-2004	(TremBLrel. 26, Last annotation update)
DE	Hypothetical protein Spel1285.	
GN	OrderedLocustNames=Spel1285;	
OS	Streptococcus pyogenes (serotype M3).	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OX	Streptococcus.	
RN	NCBI_TaxID=198466;	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=SSI-1.	
RX	MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;	
RA	Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,	
RA	Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,	
RA	Hayashi H., Hattori M., Hamada S.;	
RT	"Genome sequence of an M3 strain of Streptococcus pyogenes reveals a	
RT	large-scale genomic rearrangement in invasive strains and new insights	
RT	into phage evolution.";	
RL	Genome Res. 13:1042-1055 (2003).	
CC	-!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by	
CC	an amide bond (By similarity).	
DR	EMBL; AP005145; BAC64380.1; -.	
DR	GO; GO:0009986; C:cell surface; IEA.	
DR	GO; GO:0005618; C:cell wall; IEA.	
DR	InterPro; IPR001899; Gram_pos_anchor.	
DR	InterPro; IPR007093; LRR_Tp.	
DR	Pfam; PF00746; Gram_pos_anchor; 1.	
DR	TIGRFAMs; TIGR01167; LPXTG anchor; 1.	
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.	
KW	Cell wall; Hypothetical protein; Peptidoglycan-anchor.	
SQL	SEQUENCE 1008 AA; 111537 MW; 793C598CE094EE92 CRC64;	
	Query Match	29.1%; Score 293; DB 2; Length 1008;
	Best Local Similarity	99.3%; Pred. No. 7,8e-287;
	Matches	993; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY	1	MKGHLKTVLTLTTVVVTHNQEVPSLVKPEIPILKQTAASSISGADYAESGSKLKLINE 60
DB	1	MKGHLKTVLTLTTVVVTHNQEVPSLVKPEIPILKQTAASSISGADYAESGSKLKLINE 60
QY	61	TSGPVDDTVTDLFSDKRTTPEKIDNLAQPREQELKAVTENTESEKQITSGSQLEQSK 120
DB	61	TSGPVDDTVTDLFSDKRTTPEKIDNLAQPREQELKAVTENTESEKQITSGSQLEQSK 120
QY	121	SLSLNKTVPSSTSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHVLVPSQAADGTQLIQVAS 180
DB	121	SLSLNKTVPSSTSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHVLVPSQAADGTQLIQVAS 180
QY	181	FAPTPDKKTAIAEYTSRAGENGISQLDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
DB	181	FAPTPDKKTAIAEYTSRAGENGISQLDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
QY	241	DAFYDNKNIAEVNLPESLETISDYAFALHAKQIDLPDNLKAIGELAFDNDQITGKLSLP 300
DB	241	DAFYDNKNIAEVNLPESLETISDYAFALHAKQIDLPDNLKAIGELAFDNDQITGKLSLP 300
QY	301	QOLMLAERAPKSNHIKTIEFRGNLSKVIGEASFQNDLSQLMPLDGLKIESEAPTGNP 360
DB	301	QOLMLAERAPKSNHIKTIEFRGNLSKVIGEASFQNDLSQLMPLDGLKIESEAPTGNP 360
QY	361	GDDHNNRVVLWTKSGKNPGSLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB	361	GDDHNNRVVLWTKSGKNPGSLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
QY	421	GFNSKGLQKVRKNLKEIPKQHNGVTITEIGDGNFRNVDFQNTLRKYDLEEVKLPSTIR 480
DB	421	GFNSKGLQKVRKNLKEIPKQHNGVTITEIGDGNFRNVDFQNTLRKYDLEEVKLPSTIR 480
QY	481	KIGAFAPSNLKSFEASDDLEEKEGAFMNNRIETLEKDKLVITIGDAAFINHIYAI 540
DB	481	KIGAFAPSNLKSFEASDDLEEKEGAFMNNRIETLEKDKLVITIGDAAFINHIYAI 540
QY	541	LPESVQETIGRSAPFRONGANNLIPMGSKVKTIGEMAFISNRLEHLDLSEQKLTETIPVQAF 600

Db 541 LPESVQIEGRSAFRQNGANNLIFMGSKVKTIGEMAFNSLRLEHLDLSEKQLTETPVAQF 600
Qy 601 SDNALKEVLLPASLTKTIRREAPKQHLKQLEVASALSALSHIAFNALDDNDGDSQFQFNKVVYK 660
Db 601 SDNALKEVLLPASLTKTIRREAPKQHLKQLEVASALSALSHIAFNALDDNDGDSQFQFNKVVYK 660
Qy 661 THHNSYALADGHEHFDVDPDKLSTIVDLKILKLEGLDYSLRQTTTOTQPRDMTTAGKA 720
Db 661 THHNSYALADGHEHFDVDPDKLSTIVDLKILKLEGLDYSLRQTTTOTQPRDMTTAGKA 720
Qy 721 LLSKSNLRQGEKQFLQBAQFPLGRVDLDKAIKAKEKALVTKATKNGQLLERSINKAVL 780
Db 721 LLSKSNLRQGEKQFLQBAQFPLGRVDLDKAIKAKEKALVTKATKNGQLLERSINKAVL 780
Qy 781 AYNSAIIKANKVRLKLEKLDLTLGLVEGKGLAQATMVGQVYLLKTPPLPBYIYGLNVY 840
Db 781 AYNSAIIKANKVRLKLEKLDLTLGLVEGKGLAQATMVGQVYLLKTPPLPBYIYGLNVY 840
Qy 841 FPKSGKLIYALDMSDTIGEGQDAYGNPILNVDEDEGNYHALAVATLADYEGLDIKITILN 900
Db 841 FPKSGKLIYALDMSDTIGEGQDAYGNPILNVDEDEGNYHALAVATLADYEGLDIKITILN 900
Qy 901 SKLSQLTSTRQVPTAAHYRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGL 960
Db 901 SKLSQLTSTRQVPTAAHYRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGL 960
Qy 961 QSNPKTNRGRHSAIIPRTGSKGSFVYGILGYTSVALLSLI 1000
Db 961 QSNPKTNRGRHSAIIPRTGSKGSFVYGILGYTSVALLSLI 1000

RESULT 4
Q8K7X8
ID Q8K7X8 PRELIMINARY; PRT; 999 AA.
AC Q8K7X8;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Putative surface antigen.
GN OrderedLocusNames=SpvM3_0569;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus;
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (By similarity).
CC EMBL; AE014148; AA079176.1; -;
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR007093; LRR_Tp.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRPFAM; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Complete proteome; Peptidoglycan-anchor.
SQ SEQUENCE 999 AA; 110518 MW; 39F9801882BAEF7C CRC64;

Query Match 28.1%; Score 283; DB 2; Length 999;
Best Local Similarity 99.3%; Pred. No. 1e-276; Indels 0; Gaps 0;
Matches 983; Conservative 0; Mismatches 7;

RESULT 5
Q8E6V3
ID Q8E6V3 PRELIMINARY; PRT; 1055 AA.
AC Q8E6V3;

Qy 11 TLTVSVVTHNQEVFSLVKEPILKQTOASSISGADYAESGSKKLKINETSQVDDTVT 70
Db 2 TLTVSVVTHNQEVFSLVKEPILKQTOASSISGADYAESGSKKLKINETSQVDDTVT 61
Qy 71 DLFSDKRTTPEKIKDNIAKGPREQELKAVTENTESEKQITSGSQLEQSKESLSLNTKTVPS 130
Db 62 DLFSDKRTTPEKIKDNIAKGPREQELKAVTENTESEKQITSGSQLEQSKESLSLNTKTVPS 121
Qy 131 TSNWEICDFTKGTNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIOVASFAFPDKKTA 190
Db 122 TSNWEICDFTKGTNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIOVASFAFPDKKTA 181
Qy 191 IAEYTSRAGENGEISQLDVKGKEIINEGEVFNPSYLLAKKVTIPTGYKHIGQDAFVNDKNIA 250
Db 182 IAEYTSRAGENGEISQLDVKGKEIINEGEVFNPSYLLAKKVTIPTGYKHIGQDAFVNDKNIA 241
Qy 251 EVNLPESLETISDYAFALHALKQIDLPDNLKAIAGELAFFDNQITGKLSLPRQLMRLAERA 310
Db 242 EVNLPESLETISDYAFALHALKQIDLPDNLKAIAGELAFFDNQITGKLSLPRQLMRLAERA 301
Qy 311 FKSNIHITIEPRGNSLVKIGEASQDNDLSQMLPDGLEKIESAFTGNPGDDHNNRVV 370
Db 302 FKSNIHITIEPRGNSLVKIGEASQDNDLSQMLPDGLEKIESAFTGNPGDDHNNRVV 361
Qy 371 LMTSGKNPSGLATENTYVNPDKSLMOESPRIQVTKWLERDFTYQKNSVTGFSKGLQKV 430
Db 362 LMTSGKNPSGLATENTYVNPDKSLMOESPRIQVTKWLERDFTYQKNSVTGFSKGLQKV 421
Qy 431 KRKNLEIPKOHNGVTITEIGDNAFRNVDFONKTLRYDLEEVKLPSTIRKIGAFQSN 490
Db 422 KRKNLEIPKOHNGVTITEIGDNAFRNVDFONKTLRYDLEEVKLPSTIRKIGAFQSN 481
Qy 491 NLKSFEASDDLEETIEKGAFAFNNRIETLEKDKLVTIGDAAPHINHIYAVILPESVQIEGR 550
Db 482 NLKSFEASDDLEETIEKGAFAFNNRIETLEKDKLVTIGDAAPHINHIYAVILPESVQIEGR 541
Qy 551 SAFQNGANNLI FMGSKVKTIGEMAFNSLRLEHLDLSEKQLTETPVAQFSDNALKEVLL 610
Db 542 SAFQNGANNLI FMGSKVKTIGEMAFNSLRLEHLDLSEKQLTETPVAQFSDNALKEVLL 601
Qy 611 PASLKTIRREAFKKNHLKQLEVASALSALSHIAFNALDDNDGDSQFQFNKVVVTHHNSYALAD 670
Db 602 PASLKTIRREAFKKNHLKQLEVASALSALSHIAFNALDDNDGDSQFQFNKVVVTHHNSYALAD 661
Qy 671 GEHFIVDPDKLSSTIVDLKILKLEGLDYSLRQTTTOTQPRDMTTAGKALLSKLRQ 730
Db 662 GEHFIVDPDKLSSTIVDLKILKLEGLDYSLRQTTTOTQPRDMTTAGKALLSKLRQ 721
Qy 731 EKQKFLQBAQFPLGRVDLDKAIKAKEKALVTKATKNGQLLERSINKAVLAYNNSAIKKA 790
Db 722 EKQKFLQBAQFPLGRVDLDKAIKAKEKALVTKATKNGQLLERSINKAVLAYNNSAIKKA 781
Qy 791 NVKRLKELDLTLGLVEGKGLAQATMVGQVYLLKTPPLPBYIYGLNVYFDKSGKLIYA 850
Db 782 NVKRLKELDLTLGLVEGKGLAQATMVGQVYLLKTPPLPBYIYGLNVYFDKSGKLIYA 841
Qy 851 LDMSDTIGEGQDAYGNPILNVDEDEGNYHALAVATLADYEGLDIKITILSKLSQLTISR 910
Db 842 LDMSDTIGEGQDAYGNPILNVDEDEGNYHALAVATLADYEGLDIKITILSKLSQLTISR 901
Qy 911 QVPTAAHYRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGR 970
Db 902 QVPTAAHYRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGR 961
Qy 971 HSAILPRTGSKGSFVYGILGYTSVALLSLI 1000
Db 962 HSAILPRTGSKGSFVYGILGYTSVALLSLI 991

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein gbs0456.
 GN OrderedLocusNames=gbs0456;
 OS Streptococcus agalactiae (serotype III).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=216495;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NEM316 / Serotype III;
 RX MEDLINE=22242508; PubMed=12354221;
 RA Glaeser P., Ruzniok C., Buchrieser C., Chevalier F., Frangeul L.,
 RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
 RA Kunst F.;
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
 RT invasive neonatal disease."
 RL Mol. Microbiol. 45:1499-1513 (2002).
 DR EMBL; AL766845; CAD46100.1; -;
 DR Sagalib; gbs0456; -;
 DR InterPro; IPR007093; LRR_Tp.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 1055 AA; 117976 MW; 9DD56FBAD171BD8 CRC64;

 Query Match 8.8%; Score 89; DB 2; Length 1055;
 Best Local Similarity 100.0%; Pred. No. 3.6e-80;
 Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 805 LVGEKGLAQAATWVGVLTKTLPPLPEYIYGLNLYFVKSGKLIYALDMSDTIGSEQKDA 864
 DB 806 LVGEKGLAQAATWVGVLTKTLPPLPEYIYGLNLYFVKSGKLIYALDMSDTIGSEQKDA 865

 QY 865 YGNPILNVDEDEGHALAVATLADYEG 893
 DB 866 YGNPILNVDEDEGHALAVATLADYEG 894

 RESULT 6
 Q8E1D6 PRELIMINARY; PRT; 1055 AA.
 AC Q8E1D6;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Cell wall surface anchor family protein.
 GN OrderedLocusNames=SAG0421;
 OS Streptococcus agalactiae (serotype V).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=216466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2603 V/R / Serotype V;
 RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
 RA Tettelin H., Masignani V., Cieleszczak M.J., Eisen J.A., Peterson S.N.,
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,
 RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
 RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
 RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
 RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
 RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
 RA Fraser C.M.;
 RT "Complete genome sequence and comparative genomic analysis of an
 RT emerging human pathogen, serotype V Streptococcus agalactiae."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
 DR EMBL; AE014211; AAM99327.1; -;
 DR TIGR; SAG0421; -;
 DR InterPro; IPR007093; LRR_Tp.
 KW Complete proteome.
 SQ SEQUENCE 1055 AA; 118129 MW; 2DD8B8715B17F74E CRC64;

Query Match 7.5%; Score 76; DB 2; Length 1055;
 Best Local Similarity 100.0%; Pred. No. 5.3e-67;
 Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 805 LVGEKGLAQAATWVGVLTKTLPPLPEYIYGLNLYFVKSGKLIYALDMSDTIGSEQKDA 864
 DB 806 LVGEKGLAQAATWVGVLTKTLPPLPEYIYGLNLYFVKSGKLIYALDMSDTIGSEQKDA 865

 QY 865 YGNPILNVDEDEGHALAVATLADYEG 880
 DB 866 YGNPILNVDEDEGHALAVATLADYEG 881

 RESULT 7
 Q9ZRT5 PRELIMINARY; PRT; 245 AA.
 AC Q9ZRT5;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Glutathione transferase AtGST 10 (EC 2.5.1.18).
 GN Name=gat10; Synonym=MERF-28;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=whole plant;
 RA Dixon D.P., Cole D.J., Edwards R.;
 RT "Identification and cloning of AtGST 10 (Accession Nos. AJ131580 and
 RT AJ132398), members of a novel type of plant glutathione
 RT transferases".
 RL Plant Physiol. 119:1568-1568 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned P1 and TAC clones".
 RL DNA Res. 5:41-54 (1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the GST superfamily.
 DR EMBL; AJ131580; CAA10457.1; -;
 DR EMBL; AB010072; BAB09723.1; -;
 DR EMBL; AY054659; AAK96850.1; -;
 DR EMBL; AY072466; AAL66881.1; -;
 DR PIR; T51594; T51594.
 DR HSP; P30712; 1LJR.
 DR GO; GO:004364; F:glutathione transferase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR010987; GST_C_like.

```

DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
KW Transferrase.
SQ SEQUENCE 245 AA; 27653 MW; 6F16FE1399E793B7 CRC64;

Query Match      0.9%; Score 9; DB 2; Length 245;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAAAEQLL 937
DB 133 AAAAAEQLL 141

RESULT 8
Q9NW78 PRELIMINARY; PRT; 279 AA.
AC Q9NW78;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Hypothetical protein FLJ10244.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryo;
RX PubMed:14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirao K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togliya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshiyawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu P., Wakebe H.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama J., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 16:40-45(2004).
DR EMBL: AK001106; BAA91506.1; -.
DR GO: GO:0005085; F:guanylyl-nucleotide exchange factor activity; IEA.
DR GO: GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR InterPro: IPR001895; RasGEF CDC25.
DR InterPro: IPR008937; Ras GEF.
DR Pfam: PF00617; RasGEF; 1.
DR SMART: SM00147; RasGEF; 1.
DR PROSITE: PS50009; RASGEF CAT; 1.
DR SEQUENCE 279 AA; 31856 MW; 92195039C78F56EA CRC64;

Query Match      0.9%; Score 9; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSSES 952
DB 88 SEKSSSSSES 96

RESULT 10
Q8C134 PRELIMINARY; PRT; 523 AA.
AC Q8C134;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
DE library, clone:4732485D01 product:RAL-A EXCHANGE FACTOR RALGPS2
DE homolog.
DE Name=Raisgs2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
```

RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RA The FANTOM Consortium;
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akai H. S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AK029049; BAC26264.1; -.
 DR MGD; MGI:1925505; Ralgs2.
 DR GO; GO:0005085; F:guanylyl-nucleotide exchange factor activity; IEA.
 DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
 DR InterPro; IPR001895; RasGRF CDC25.
 DR InterPro; IPR008937; Ras_GEF.
 DR Pfam; PF00617; RasGEF; 1.
 DR SMART; SM00147; RasGEF; 1.
 DR PROSITE; PS50009; RasGEF CAT; 1.
 SQ SEQUENCE 523 AA; 57863 MW; D89582752DBA9F4D CRC64;
 Query Match 0.9%; Score 9; DB 2; Length 523;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 944 SEKSSSES 952
 DB |||||
 18 SEKSSSES 26
 RESULT 11
 Q80YA6 PRELIMINARY; PRT; 564 AA.
 ID Q80YA6
 AC Q80YA6
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DE Mus musculus adult male cecum cDNA, RIKEN full-length enriched
 DE library, clone:9130014M22 product:RAL-A EXCHANGE FACTOR RALGFS2
 DE homolog.
 GN Name=4921528G01Rik;

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Ralgs2 protein.
 GN Name=Ralgs2;
 OS Mus musculus (Mouse);
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Roha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBSJ databases.
 CC -!- SIMILARITY: Contains 1 PH domain.
 DR EMBL; BC043132; AAH43132.1; -.
 DR MGD; MGI:1925505; Ralgs2.
 DR GO; GO:0005085; F:guanylyl-nucleotide exchange factor activity; IEA.
 DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR011036; PH related.
 DR InterPro; IPR001895; RasGRF CDC25.
 DR InterPro; IPR008937; Ras_GEF.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00617; RasGEF; 1.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00147; RasGEF; 1.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 DR PROSITE; PS50009; RasGEF CAT; 1.
 SQ SEQUENCE 564 AA; 62524 MW; A2988CD76323F0EE CRC64;
 Query Match 0.9%; Score 9; DB 2; Length 564;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 944 SEKSSSES 952
 DB |||||
 18 SEKSSSES 26
 RESULT 12
 Q9D2Y7 PRELIMINARY; PRT; 568 AA.
 ID Q9D2Y7
 AC Q9D2Y7
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult male cecum cDNA, RIKEN full-length enriched
 DE library, clone:9130014M22 product:RAL-A EXCHANGE FACTOR RALGFS2
 DE homolog.
 GN Name=4921528G01Rik;


```

Db          18 SEKSSSES 26
RESULT 14
Q9FHE1      PRELIMINARY;      PRT;      590 AA.
AC Q9FHE1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glutathione transferase-like.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eusoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20181125; PubMed=10718197;
RX Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:31-63(2000).
CC -1- SIMILARITY: Contains 1 Myb-like domain.
DR EMBL; AB019225; BAB11098.1; -.
DR HSP; Q9ZVQ3; 1E6B.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_Clike.
DR InterPro; IPR004045; GST_Nterm.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS50090; MYB_3; 1.
KW Transferase.
SQ SEQUENCE 590 AA; 67554 MW; E85F66525F6C7AF5 CRC64;

Query Match          0.9%; Score 9; DB 2; Length 590;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAAEAFQLL 937
Db 132 AAAAEAFQLL 140

RESULT 15
Q7TP26      PRELIMINARY;      PRT;      590 AA.
AC Q7TP26;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ralgs2 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Egg;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

```

```

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; BC052663; AAHS2663.1; -.
DR GO; GO:0005085; F:guanylyl-nucleotide exchange factor activity; IEA.
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR InterPro; IPR011849; PH.
DR InterPro; IPR011036; PH-related.
DR InterPro; IPR001895; RasGRF_CDC25.
DR InterPro; IPR008937; Ras_GEF.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00617; RasGEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00147; RasGEF; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50009; RasGEF_CAT; 1.
SQ SEQUENCE 590 AA; 65458 MW; CAC5CD67FF04DE0F CRC64;

Query Match          0.9%; Score 9; DB 2; Length 590;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952
Db 18 SEKSSSES 26

RESULT 16
Q8BZU2      PRELIMINARY;      PRT;      590 AA.
AC Q8BZU2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus adult male colon cDNA, RIKEN full-length enriched
DE library, clone:9030616022 product:RAL-A EXCHANGE FACTOR RALGPS2
DE homolog.
GN Name=Ralgs2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;

```



```

RN [4]
RP SEQUENCE FROM N.A.
RA Wilson R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AC004884; AAS00352.1; -.
KW Hypothetical protein.
FT NON_TER 12 12
SQ SEQUENCE 12 AA; 1271 MW; 98B16D259CB33B05 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSE 951
DB 5 SEKSSSSE 12
|||||

RESULT 20
Q982N8 PRELIMINARY; PRT; 67 AA.
AC Q982N8;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Mr8560 protein.
GN OrderedLocusNames=mr8560;
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideawara Y., Iehikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
DL DNA Res. 7:331-338(2000).
DR EMBL; AP03014; BAB54418.1; -.
KW Complete proteome.
SQ SEQUENCE 67 AA; 7250 MW; 27F1924C9BFC93B5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 67;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 757 KALVTKKA 764
DB 52 KALVTKKA 59
|||||

RESULT 21
Q7UVJ9 PRELIMINARY; PRT; 98 AA.
AC Q7UVJ9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Probable ethanolamine utilization protein RutM.
GN Name=eutM; OrderedLocusNames=RB2586;
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1;
RA MEDLINE=22735913; PubMed=12835416; DOI=10.1073/pnas.1431443100;

RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294137; CAD72725.1; -.
DR InterPro; IPR000249; Bact_microcomp.
DR Pfam; PF00936; Bact_microcomp; 1.
DR ProDom; PD003442; Bact_microcomp; 1.
KW Complete proteome.
SQ SEQUENCE 98 AA; 9953 MW; 60966543693CC8C2 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 687 DLEKILKL 694
DB 85 DLEKILKL 92
|||||

RESULT 22
RL3E_SULTO
ID _RL3E_SULTO STANDARD; PRT; 106 AA.
AC P58376;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE 50S ribosomal protein L30e.
GN Name=rlp30e; OrderedLocusNames=ST0274;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OX Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JCM 10545 / 7;
MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7."
DL DNA Res. 8:123-140(2001).
CC -1- SIMILARITY: Belongs to the ribosomal protein L30e family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AP000982; BAB65241.1; -.
DR HSP; P29160; 1H7M.
DR HAMAP; MF_00481; -.
DR InterPro; IPR000231; Ribosomal_L30e.
DR InterPro; IPR004038; Ribosomal_L7A.
DR Pfam; PF01248; Ribosomal_L7Ae; 1.
DR ProDom; PD004495; Ribosomal_L7A; 1.
DR PROSITE; PS00709; RIBOSOMAL_L30E_1; 1.
DR PROSITE; PS00993; RIBOSOMAL_L30E_2; 1.
KW Complete proteome; Ribosomal protein.
SQ SEQUENCE 106 AA; 11616 MW; DB9F4D86E5CF57CEF CRC64;

Query Match 0.8%; Score 8; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 501 LBEIKEGA 508
DB 98 LBEIKEGA 105

RESULT 23
QNSZ8
ID Q9NSZ8 PRELIMINARY; PRT; 116 AA.
AC Q9NSZ8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein DKFZp434H0512 (Fragment).
GN Name=DKFZp434H0512;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Duesterhoeft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137620; CAB70846.1; -.
DR PIR; T46320; T46320.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 116 AA; 13372 MW; CED738D670665BE2 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 116;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 930 ABAEQOLL 937
DB 83 ABAEQOLL 90

RESULT 24
Q87WF0
ID Q87WF0 PRELIMINARY; PRT; 165 AA.
AC Q87WF0;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Transcription elongation factor GreB.
GN Name=greB-2; OrderedLocNames=PSPT04601;
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA Buell C.R., Joardar V., Lindeberg M., Selenkut J., Paulsen I.T.,
RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,
RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
RA Usterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA Bender C.L., White O., Fraser C.M., Collmer A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
Pseudomonas syringae pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
CC -!- FUNCTION: Necessary for efficient RNA polymerase transcription
elongation past template-encoded arresting sites. The arresting
sites in DNA have the property of trapping a certain fraction of
elongating RNA polymerases that pass through, resulting in locked
CC ternary complexes. Cleavage of the nascent transcript by cleavage

```

```

CC factors such as greA or greB allows the resumption of elongation
CC from the new 3'-terminus (By similarity).
CC -!- SIMILARITY: Belongs to the greA/greB family.
DR EMBL; AE016872; AA058047.1; -.
DR HSSP; P21346; 1GRJ.
DR TIGR; PSPT04601; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003711; F:transcriptional elongation regulator activity; IEA.
DR GO; GO:0003746; P:translation elongation factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001437; GreA_GreB.
DR Pfam; PF01272; GreA_GreB_1.
DR ProDom; PD004918; GreA_GreB; 1.
KW Complete proteome; DNA-binding; Elongation factor; Transcription;
KW Transcription regulation.
SQ SEQUENCE 165 AA; 18080 MW; 7A47CC37FD178B58 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 165;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 928 NAAAEAEQ 935
DB 10 NAAAEAEQ 17

RESULT 25
Q9SB23
ID Q9SB23 PRELIMINARY; PRT; 174 AA.
AC Q9SB23;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Beta-amylase (EC 3.2.1.2) (Fragment).
GN Name=beta-Amy;
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Haruna Nijo;
RA Okada Y., Kihara M., Kuroda H., Yoshigi N., Ito K.;
RT "Cloning and Sequencing of the Promoter Region of the Seed Specific
beta-Amylase Gene from Barley."
RL J. Plant Physiol. 156:762-767(2000).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
polysaccharides so as to remove successive maltose units from the
non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; D63574; BAA09793.1; -.
DR HSSP; P16098; 1BY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0002272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001554; Glyco_hydro_14.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
FT NON_TER
SQ SEQUENCE 174 AA; 19593 MW; C8AFA61737153279 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62

```

```

Matches      8;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy      794 RLEKELDL 801
Db      44 RLEKELDL 51

RESULT 28
Q7T3A2
ID Q7T3A2 PRELIMINARY; PRT; 241 AA.
AC Q7T3A2;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Zgc:64012.
GN ORFNames=zgc:64012;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uebin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faneq J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053197; AAH53197.1; -.
DR ZFIN; ZDB-GENE-040426-2639; zgc:64012.
DR GO; GO:0005525; P:GTP binding; IEA.
SQ SEQUENCE 241 AA; 26617 MW; C625FE64088D3D98 CRC64;

Query Match      0.8%; Score 8; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches      8;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy      349 EKIESEAF 356
Db      198 EKIESEAF 205

RESULT 29
Q7NST9
ID Q7NST9 PRELIMINARY; PRT; 242 AA.
AC Q7NST9;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Probable ABC transporter, ATP-binding protein.
GN OrderedLocusNames=CV3331;

Matches      8;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy      948 SSSSANS 955
Db      42 SSSSANS 49

RESULT 27
Q8TW63
ID Q8TW63 PRELIMINARY; PRT; 240 AA.
AC Q8TW63;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Uncharacterized protein specific for M.kandleri, MK-27 family.
GN OrderedLocusNames=Mk1173;
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus;
OX NCBI_TaxID=23320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014; DOI=10.1073/pnas.032671499;
RA Slesarev A.I., Mezhevaeva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AE010408; AA002386.1; -.
KW Complete proteome.
SQ SEQUENCE 240 AA; 25400 MW; 59813D9792655BEC CRC64;

Query Match      0.8%; Score 8; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches      8;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy      948 SSSSANS 955
Db      42 SSSSANS 49

RESULT 26
Q9FR02
ID Q9FR02 PRELIMINARY; PRT; 237 AA.
AC Q9FR02;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE AV-9/Cf-9 rapidly elicited protein 1.
GN Name=ACRE1;
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20312759; PubMed=10852940;
RA Durrant W.B., Rowland O., Piedras P., Hammond-Kosack K.E., Jones J.D.;
RT "cDNA-AFLP reveals a striking overlap in race-specific resistance and
RT wound response gene expression profiles."
RL Plant Cell 12:963-977(2000).
DR EMBL; AF211527; AAG43545.1; -.
DR HSP; O80337; 2GCC.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003709; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001471; TF_ERF.
DR PRINTS; PR00367; ETRHSPLEWNT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
SQ SEQUENCE 237 AA; 26513 MW; 823E5154327BB87 CRC64;

Query Match      0.8%; Score 8; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches      8;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy      948 SSSSANS 955
Db      42 SSSSANS 49

```

OS Chromobacterium violaceum.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Chromobacterium.
 OX NCBI_TaxID=536;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12472 / DSM 30191;
 RX MEDLINE=2282880; PubMed=14500782; DOI=10.1073/pnas.1032124100;
 RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
 RA Antonio R.V., Almeida F.C. de Almeida L.G.P., de Almeida R.,
 RA Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
 RA Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
 RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
 RA Bordignon J., Brígido M.M., Brito C.A., Brocchi M., Burity H.A.,
 RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
 RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
 RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Faicao C.L.,
 RA Faninatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
 RA Gazzino R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
 RA Grattapaglia D., Griesard E.C., Hanna E.S., Jardim S.N., Laurino J.,
 RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
 RA Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,
 RA di Mauro S.M.Z., de Medeiros S.R.B., Meisner R.V., Moreira M.A.M.,
 RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
 RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
 RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
 RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
 RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seunaez H.N.,
 RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
 RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
 RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
 RA Vettore A., Wasseem R., Zaha A., Simpson A.J.G.;
 RT "The complete genome sequence of Chromobacterium violaceum reveals
 RT remarkable and exploitable bacterial adaptability."
 RT Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
 RC -!- SIMILARITY: Belongs to the ABC transporter family.
 DR EMBL; AB016921; BAA06095.1; -.
 DR HSP; O58663; LG9X.
 DR GO; GO:0016020; Cmembrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
 DR GO; GO:0042829; P:defense response to pathogen; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR000767; Disease_reisist.
 DR Pfam; PF00005; ABC_tran; 1.
 DR PRINTS; PR00364; DISEASERSIST.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00189; LIPOYL; UNKNOWN_1.
 KW ATP-binding; Complete proteome.
 SQ SEQUENCE 242 AA; 26822 MW; 912D717124132EA8 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 242;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 795 LEKELDLL 802
 Db 115 LEKELDLL 122
 RESULT 30
 P93664 PRELIMINARY; PRT; 253 AA.
 AC P93664;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE LI818r-1 protein precursor.

GN Name=LI818r-1;
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CC620;
 RA Guertin M.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X95326; CAA64632.1; -.
 DR PIR; T08175; T08175.
 DR GO; GO:0009785; P:photosynthesis light harvesting; IEA.
 DR InterPro; IPR001344; Chloro Abbind.
 DR Pfam; PF00504; Chloroa_b_bind; 1.
 KW Signal.
 FT SIGNAL 1 32 Potential.
 FT CHAIN 33 253 Potential.
 SQ SEQUENCE 253 AA; 27556 MW; 985274D58C5BBF1D CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 253;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 276 LPDNLKAI 283
 Db 246 LPDNLKAI 253
 RESULT 31
 Q9AB78 PRELIMINARY; PRT; 255 AA.
 AC Q9AB78;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DE Hypothetical protein.
 DE OrderedLocustNames=BT0232;
 OS Bacteroides thetaiotaomicron.
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Bacteroidaceae; Bacteroides.
 OX NCBI_TaxID=818;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPI-5482 / ATCC 29148;
 RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.10800029;
 RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
 RA Chiang H.C., Hooper L.V., Gordon J.I.;
 RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
 RL Science 299:2074-2076(2003).
 DR EMBL; AB016926; BAA075339.1; -.
 KW Complete proteome.
 SQ SEQUENCE 255 AA; 29434 MW; 4894A4FCS168A8E2 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 255;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 696 EGLDYSTL 703
 Db 215 EGLDYSTL 222
 RESULT 32
 Q8UYE5 PRELIMINARY; PRT; 255 AA.
 AC Q8UYE5;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative chemokine binding protein.
 OS Camelpox virus (strain M-96).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

```

OC Orthopoxvirus.
OX NCBI_TaxID=203173;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22031204; PubMed=12033760; DOI=10.1006/viro.2001.1343;
RX Alfonso C.L., Tulman E.R., Lu Z., Zsak L., Sandybaev N.T.,
RA Kerembekova U.Z., Zaitsev V.L., Kutish G.F., Rock D.L.;
RT "The genome of camelpox virus.";
RL Virology 295:1-9(2002).
DR EMBL; AF438165; AAL73919.1; -.
DR EMBL; AF438165; AAL73919.1; -.
DR HSSP; O73568; 1CQ3.
DR Pfam; PF02250; Orthopox_35kd; 1.
DR PIRSF; PIRSF003696; VAC_C23L; 1.
SQ SEQUENCE 255 AA; 27659 MW; 84B8DF0C01CE31FE CRC64;

Query Match 0.8%; Score 8; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 VTENTESE 106
Db 61 VTENTESE 68

RESULT 33
ID Q6YLZ5 PRELIMINARY; PRT; 255 AA.
AC Q6YLZ5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Secreted chemokine binding protein.
OS Camelpox virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=28873;
RN [1]
RP SEQUENCE FROM N.A.
RA Saudi;
RC STRAIN=CP-1, CP-1231 Haut Dubai, CP-1260/95 Haut Dubai, CP-5, and
RA Mikheev M.V., Feshchenko M.V., Shchelkunov S.N.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY102980; AAM76329.1; -.
DR EMBL; AY102981; AAM76330.1; -.
DR EMBL; AY102982; AAM76331.1; -.
DR EMBL; AY102984; AAM76333.1; -.
DR EMBL; AY102983; AAM76332.1; -.
DR EMBL; AY102983; AAM76332.1; -.
DR InterPro; IPR003184; Orthopox_35kd.
DR Pfam; PF02250; Orthopox_35kd; 1.
DR PIRSF; PIRSF003696; VAC_C23L; 1.
SQ SEQUENCE 255 AA; 27659 MW; 84B8DF0C01CE31FE CRC64;

Query Match 0.8%; Score 8; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 VTENTESE 106
Db 61 VTENTESE 68

RESULT 34
ID Q776C0 PRELIMINARY; PRT; 255 AA.
AC Q776C0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE CMF1L (CMP206R).
GN Name=CMF1L; Synonyms=CMF206R;
OS Camelpox virus (strain CMS).

```

```

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=203172;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CMS;
RX PubMed=11907336;
RA Gubser C., Smith G.L.;
RT "The sequence of camelpox virus shows it is most closely related to
RL variola virus, the cause of smallpox.";
RL J. Gen. Virol. 83:855-872(2002).
DR EMBL; AY009089; AAG37455.1; -.
DR EMBL; AY009089; AAG37455.1; -.
DR InterPro; IPR003184; Orthopox_35kd.
DR InterPro; IPR009173; VAC_C23L.
DR Pfam; PF02250; Orthopox_35kd; 1.
DR PIRSF; PIRSF003696; VAC_C23L; 1.
SQ SEQUENCE 255 AA; 27659 MW; 84B8DF0C01CE31FE CRC64;

Query Match 0.8%; Score 8; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 VTENTESE 106
Db 61 VTENTESE 68

RESULT 35
ID Q9H0J8 PRELIMINARY; PRT; 268 AA.
AC Q9H0J8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein DKFZp434B0516.
GN Name=DKFZp434B0516;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Kohrer K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL136767; CAB66701.1; -.
KW Hypothetical protein.
SQ SEQUENCE 268 AA; 31293 MW; BBE2BA96541635D7 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 793 KRLSKELD 800
Db 83 KRLSKELD 90

RESULT 36
ID O96191 PRELIMINARY; PRT; 284 AA.
AC O96191;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PFB0490c.
GN Name=PFB0490c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=99021743; PubMed=9804551; DOI=10.1126/science.282.5391.1126;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalloom S., Mason T., Yu K., Fujii C., Pedersen J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perte A.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shalloom S.J., Sub B., Peterson J., Angiuoli S.,
RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE001398; AAC71887.1; -.
DR PIR; E71613; E71613.
KW Hypothetical protein.
SQ SEQUENCE 284 AA; 32541 MW; 41B35EFBC2C82B43 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 284;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 EVNLPESL 258
DB 269 EVNLPESL 276
|||||

RESULT 37
Q926G6 PRELIMINARY; PRT; 296 AA.
AC Q926G6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative sugar uptake ABC transporter periplasmic solute-binding
DE protein.
GN ORFNames=SWB21377;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431; DOI=10.1073/pnas.161294698;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoevelter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,693-kb pSymB megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL591985; CAC49355.1; -.
DR PIR; C95961; C95961.
DR HSSP; P02925; 2DRI.
DR Pfam; PF00532; Peripla_BP_1.1.
KW Complete proteome; Plasmid.
SQ SEQUENCE 296 AA; 30625 MW; 43CB86F44829C9F3 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 751 AIAKAEKA 758
DB 98 AIAKAEKA 105
|||||

RESULT 38
Q9ABZ6 PRELIMINARY; PRT; 304 AA.
AC Q9ABZ6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein CC0074.
GN OrderedLocustNames=CC0074;
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647; DOI=10.1073/pnas.061029298;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H.M., Shetty J.,
RA Berry K.J., Utterback T.R., Tran K., Wolf A.M., Vamathevan J.J.,
RA Emolaeva M.D., White O., Salzberg S.L., Venter J.C., Shapiro L.,
RA Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005681; AAK22061.1; -.
DR PIR; A87258; A87258.
DR TIGR; CC0074; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 304 AA; 31526 MW; ED0A35B28AD493E1 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 881 ALAVATLA 888
DB 131 ALAVATLA 138
|||||

RESULT 39
Q88U35 PRELIMINARY; PRT; 316 AA.
AC Q88U35;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein lp.2713.
GN OrderedLocustNames=lp.2713;
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12565656; DOI=10.1073/pnas.0337704100;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935260; CAD64951.1; -.

```

```

KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 316 AA; 34797 MW; E9DB42A1AF03776D CRC64;

Query Match      0.8%; Score 8; DB 2; Length 316;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 LEVASALS 637
Db 227 LEVASALS 234
|||||

RESULT 40
YDZA SCHPO
ID YDZA SCHPO STANDARD; PRT; 329 AA.
AC 013717;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hypothetical protein C14C4.10c in chromosome I.
GN ORFName=SPAC14C4.10c;
OS Schizosaccharomyces pombe (Pission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feitwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squires R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vansreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Frizic C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Chuzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z98596; CAB11203.1; --
CC PIR; T37695; T37695.
CC GenDB_SPombe; SPAC14C4.10c; --
CC InterPro; IPR000086; NUDIX_hydrolase.
CC Pfam; PF00293; NUDIX; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 123 143
Db 143
Potential.

```

```

FT TRANSMEM 227 247 Potential.
FT TRANSMEM 303 323 Potential.
SQ SEQUENCE 329 AA; 38293 MW; 5DB9C067BA1E1743 CRC64;

Query Match      0.8%; Score 8; DB 1; Length 329;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 798 ELDLLTGL 805
Db 4 ELDLLTGL 11
|||||

RESULT 41
O99927
ID O99927 PRELIMINARY; PRT; 348 AA.
AC O99927;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE NADH dehydrogenase subunit 2.
GN Name=ND2;
OS Cyprinella lutrensis (Red shiner).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinella.
OX NCBI_TaxID=28791;
RN [1]
RP SEQUENCE FROM N.A.
RA Broughton R.E., Gold J.R.;
RT "Phylogenetic relationships in the North American cyprinid genus
RT Cyprinella (Actinopterygii: Cyprinidae) based on sequences of the
RT mitochondrial ND2 and ND4L genes.";
RL Copeia 2000:1-10(2000).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (By similarity).
CC -1- SIMILARITY: Belongs to the complex I subunit 2 family.
CC EMBL; AF111211; AAD13235.1; --
CC GO; GO:0005739; C:mitochondrion; IEA.
CC GO; GO:0008137; P:NADH dehydrogenase (ubiquinone) activity; IEA.
CC GO; GO:0016491; F:oxidoreductase activity; IEA.
CC GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.
CC GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
CC InterPro; IPR003917; NADH dehydrogenase.
CC InterPro; IPR010933; NADH dehydrogenase.
CC Pfam; PF06444; NADH dehydrogenase.
CC Pfam; PF00361; Oxidored_q1_1.
CC PRINTS; PR01436; NADH dehydrogenase.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 348 AA; 37586 MW; 377C3A14A3E50625 CRC64;

Query Match      0.8%; Score 8; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 LDLLTGLV 806
Db 122 LDLLTGLV 129
|||||

RESULT 42
O99944
ID O99944 PRELIMINARY; PRT; 348 AA.
AC O99944;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE NADH dehydrogenase subunit 2.
GN Name=ND2;
OS Erimonax monachus.

```


OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
 OC Cyprinidae; Erimonax.
 OX NCBI_TaxID=87719;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Broughton R.E., Gold J.R.;
 RT "Phylogenetic relationships in the North American cyprinid genus
 RT *Cyprinella* (Actinopterygii: Cyprinidae) based on sequences of the
 RT mitochondrial ND2 and ND4L genes.";
 RL Copeia 2000:1-10(2000).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (By similarity).
 CC -1- SIMILARITY: Belongs to the complex I subunit 2 family.
 DR EMBL, AF111228; AD13252.1; -.
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . . ; IEA.
 DR InterPro; IPR003917; NADH_oxred2.
 DR InterPro; IPR010933; NADH_dchy_S2_C.
 DR InterPro; IPR001750; Oxidored_q1.
 DR Pfam; PF06444; NADH_dchy_S2_C; 1.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR PRINTS; PR01436; NADHDHGNASE2.
 KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
 SQ SEQUENCE 348 AA; 37634 MW; 3417A690E3F4B9A8 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 348;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 LDLLTGLV 806
 |||||
 DB 122 LDLLTGLV 129

RESULT 43

Q8M518 PRELIMINARY; PRT; 348 AA.
 AC Q8M518;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE NADH dehydrogenase subunit 2.
 GN Name=ND2;
 OS Pelvicachromis pulcher (Rainbow krib).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
 OC Cichlidae; Pelvicachromis.
 OX NCBI_TaxID=28827;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Klett V., Meyer A.;
 RT "What, if anything, is a Tilapia?—mitochondrial ND2 phylogeny of
 RT tilapines and the evolution of parental care systems in the African
 RT cichlid fishes";
 RL Mol. Biol. Evol. 19:865-883(2002).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (By similarity).
 CC -1- SIMILARITY: Belongs to the complex I subunit 2 family.
 DR EMBL, AF317271; AM20784.1; -.
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.

DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . . ; IEA.
 DR InterPro; IPR003917; NADH_oxred2.
 DR InterPro; IPR010933; NADH_dchy_S2_C.
 DR InterPro; IPR001750; Oxidored_q1.
 DR Pfam; PF06444; NADH_dchy_S2_C; 1.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR PRINTS; PR01436; NADHDHGNASE2.
 KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
 SQ SEQUENCE 348 AA; 37879 MW; 85509DE660A1906 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 348;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 LDLLTGLV 806
 |||||
 DB 122 LDLLTGLV 129

RESULT 44

Q8HLA4 PRELIMINARY; PRT; 348 AA.
 AC Q8HLA4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE NADH dehydrogenase subunit 2.
 GN Name=ND2;
 OS Indostomus paradoxus (Armoured stickleback).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Gasterosteiformes;
 OC Indostomidae; Indostomus.
 OX NCBI_TaxID=181450;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22359432; PubMed=12470944; DOI=10.1016/S1055-7903(02)00332-9;
 RA Miya M., Takeshima H., Endo H., Ishiguro N.B., Inoue J.G., Mukai T.,
 RA Sato T.P., Yamaguchi M., Kawaguchi A., Mabuchi K., Shirai S.M.,
 RA Nishida M.;
 RT "Major patterns of higher teleostean phylogenies: a new perspective
 RT based on 100 complete mitochondrial DNA sequences.";
 RL Mol. Phylogenet. Evol. 26:121-138(2003).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (By similarity).
 CC -1- SIMILARITY: Belongs to the complex I subunit 2 family.
 DR EMBL, AP004438; BAC23594.1; -.
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . . ; IEA.
 DR InterPro; IPR003917; NADH_oxred2.
 DR InterPro; IPR010933; NADH_dchy_S2_C.
 DR InterPro; IPR001750; Oxidored_q1.
 DR Pfam; PF06444; NADH_dchy_S2_C; 1.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR PRINTS; PR01436; NADHDHGNASE2.
 KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
 SQ SEQUENCE 348 AA; 37834 MW; 3E4561B5EB4B3631 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 348;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 LDLLTGLV 806
 |||||
 DB 122 LDLLTGLV 129

RESULT 45

Q26490
ID Q26490 PRELIMINARY; PRT; 351 AA.
AC Q26490
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Retinol dehydratase.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96279157; PubMed=8663216; DOI=10.1074/jbc.271.27.16135;
RA Grun F., Noy N., Hammerling U., Buck J.;
RT "Purification, cloning, and bacterial expression of retinol
RT dehydratase from Spodoptera frugiperda";
RL J. Biol. Chem. 271:16135-16138(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95074030; PubMed=7982943;
RA Zheng Y., Bergold A., Duffel M.W.;
RT "Affinity labeling of aryl sulfotransferase IV. Identification of a
RT peptide sequence at the binding site for 3'-phosphoadenosine-5'-
RT phosphosulfate";
RL J. Biol. Chem. 269:30313-30319(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95071451; PubMed=7980593;
RA Komatsu K., Driscoll W.J., Koh Y.C., Strott C.A.;
RT "A P-loop related motif (GxxGxxK) highly conserved in
RT sulfotransferases is required for binding the activated sulfate
RT donor";
RL Biochem. Biophys. Res. Commun. 204:1178-1185(1994).
DR EMBL; U28654; AAC47136.1; -
DR PDB; 1FMJ; X-ray; A/B=1-351.
DR PDB; 1FML; X-ray; A/B=1-351.
DR GO; GO:0008146; P:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer 1; 1.
SQ SEQUENCE 351 AA; 41562 MW; 004D9CE41856FB6E CRC64;

Query Match 0.8%; Score 8; DB 2; Length 351;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 668 LADGEHFI 675
DB 304 LADGEHFI 311

RESULT 46
Q8RPN6 PRELIMINARY; PRT; 360 AA.
ID Q8RPN6
AC Q8RPN6
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CcmH.
GN Name=ccmH;
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=130b;
RX MEDLINE=21893094; PubMed=11895946;
RX DOI=10.1128/IAI.70.4.1842-1852.2002;
RA Viswanathan V.K., Kurtz S., Pedersen L.L., Abu Kwaik Y., Krcmarik K.,
RA Mody S., Cianciotto N.P.;
RT "The cytochrome c maturation locus of Legionella pneumophila promotes

iron assimilation and intracellular infection and contains a strain-
specific insertion sequence element.";
RL Infect. Immun. 70:1842-1852(2002).
DR EMBL; AF386079; AA000399.1; -
DR InterPro; IPR005616; CcmH.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03918; CcmH; 1.
DR ProDom; PD005662; CcmH; 1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
DR PROSITE; PS50005; TPR; 2.
DR PROSITE; PS50293; TPR_REGION; 1.
SQ SEQUENCE 360 AA; 41845 MW; 54703067E43171B6 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 KAIKAEK 757
DB 342 KAIKAEK 349

RESULT 47
Q9F7S2 PRELIMINARY; PRT; 360 AA.
ID Q9F7S2
AC Q9F7S2
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Predicted N-acetylmuramoyl-L-alanine amidase.
OS Gamma-proteobacterium EBAC31A08.
OC Bacteria; Proteobacteria; Gammaproteobacteria; environmental samples.
OX NCBI_TaxID=133804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20446260; PubMed=10989064; DOI=10.1126/science.289.5486.1902;
RA Beja O., Aravind L., Koonin E.V., Suzuki M.T., Hadd A., Nguyen L.P.,
RA Jovanovich S.B., Gates C.M., Feldman R.A., DeLong E.F.;
RA Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
RT "Bacterial rhodopsin: evidence for a new type of phototrophy in the
RT sea.";
RL Science 289:1902-1906(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Beja O., Aravind L., Koonin E.V., Suzuki M.T., Hadd A., Nguyen L.P.,
RA Jovanovich S.B., Gates C.M., Feldman R.A., DeLong E.F.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF279106; AAG10447.1; -
DR GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro; IPR002508; Amidase_3_hydro.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01520; Amidase_3; 1.
DR Pfam; PF01476; LysM; 1.
DR SMART; SM00646; Ami_3; 1.
DR SMART; SM00257; LysM; 1.
SQ SEQUENCE 360 AA; 40939 MW; 3313E14B8C6F3C24 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 KVKENKNL 436
DB 61 KVKENKNL 68

RESULT 48
Q87S80 PRELIMINARY; PRT; 370 AA.
ID Q87S80

```

AC Q87580;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein VP0544.
GN OrderedLocusNames=VP0544;
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739; DOI=10.1016/S0140-6736(03)12659-1;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
DR EMBL; AP005074; BAC58807.1; -.
DR InterPro; IPR005338; UPF0075.
DR Pfam; PF03702; UPF0075; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 370 AA; 43036 MW; DCSE9DF79049C16E CRC64;

Query Match 0.8%; Score 8; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

QY 279 NLKAIGEL 286
DB 59 NLKAIGEL 66

RESULT 49
Q7QF26
ID Q7QF26 PRELIMINARY; PRT; 385 AA.
AC Q7QF26;
DT 01-WAR-2004 (TrEMBLrel. 26, Created)
DT 01-WAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AgCP13368 (fragment).
GN Name=agCG54736; ORFNames=ENSGG00000009938;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008846; EAA06711.1; -.
FT NON_TER 1
SQ SEQUENCE 385 AA; 44555 MW; 3FF4E2A3FB3ABFEE CRC64;

Query Match 0.8%; Score 8; DB 2; Length 385;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 LEIKDKLV 524
DB 211 LEIKDKLV 218

RESULT 50
Q63C90
ID Q63C90 PRELIMINARY; PRT; 391 AA.
AC Q63C90;

```

```

DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Alanine racemase (EC 5.1.1.1).
GN Name=dal; ORFNames=BTZK1884;
OS Bacillus cereus ZK.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=288681;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZK;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RT "Complete genome sequence of Bacillus cereus ZK.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000001; AAU18372.1; -.
KW Isomerase.
SQ SEQUENCE 391 AA; 43417 MW; 59CD94ACF3851BD5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 988 ILGYTVA 995
DB 86 ILGYTVA 93

RESULT 51
Q739H8
ID Q739H8 PRELIMINARY; PRT; 391 AA.
AC Q739H8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Alanine racemase (EC 5.1.1.1).
GN Name=alr; OrderedLocusNames=BCE2163;
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714; DOI=10.1093/nar/gkh258;
RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT adaptations and a large plasmid related to Bacillus anthracis pX01.";
RL Nucleic Acids Res. 32:977-988(2004).
CC -! CATALYTIC ACTIVITY: L-alanine = D-alanine.
CC -! COFACTOR: Pyridoxal phosphate (By similarity).
CC -! SIMILARITY: Belongs to the alanine racemase family.
DR EMBL; AE017271; AAS41084.1; -.
DR HSSP; P10724; 1BD0.
DR TIGR; BCE2163; -.
DR GO; GO:0008784; F:alanine racemase activity; IEA.
DR GO; GO:0030170; F:pyridoxal phosphate binding; IEA.
DR GO; GO:0006522; P:alanine metabolism; IEA.
DR InterPro; IPR000821; Ala_racemase.
DR InterPro; IPR011079; Ala_racemase_C.
DR InterPro; IPR001608; Ala_racemase_N.
DR InterPro; IPR009006; Racem_decarbox_C.
DR Pfam; PF00842; Ala_racemase_C; 1.
DR Pfam; PF01168; Ala_racemase_N; 1.
DR PRINTS; PR00992; ALARACEMASE.
DR TIGRFAMs; TIGR00492; alr; 1.
DR PROSITE; PS00395; ALANINE_RACEMASE; UNKNOWN 1.
KW Complete proteome; Isomerase; Pyridoxal phosphate.
SQ SEQUENCE 391 AA; 43385 MW; .345588619898987 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;

```

```
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 988 ILGYTSVA 995
Db 86 ILGYTSVA 93
|||||
RESULT 52
Q81RG8 PRELIMINARY; PRT; 391 AA.
AC Q81RG8; Q6HZP3; Q6KTN0;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Alanine racemase.
GN Name=dal-2; OrderedLocusNames=BA2079, BAS1932, GBAA2079;
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate Porton;
RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;
RA Read T.D., Peterson S.N., Toukase N.J., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R.T., Redmond C., Thwaitte J.E., White O.,
RA Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M.,
RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Sterne;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBSJ databases.
CC -!- CATALYTIC ACTIVITY: L-alanine = D-alanine.
CC -!- COPACTORY: Pyridoxal phosphate (By similarity).
CC -!- SIMILARITY: Belongs to the alanine racemase family.
DR EMBL; AE017030; AAP25965.1; -
DR EMBL; AE017334; AAT31196.1; -
DR HSSP; P10724; INIU.
DR TIGR; BA2079; -.
DR TIGR; GBAA2079; -.
DR GO; GO:0008784; F:alanine racemase activity; IEA.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR GO; GO:0030170; F:pyridoxal phosphate binding; IEA.
DR GO; GO:0006522; P:alanine metabolism; IEA.
DR InterPro; IPR000821; Ala_racemase.
DR InterPro; IPR011079; Ala_racemase_C.
DR InterPro; IPR001608; Ala_racemase_N.
DR InterPro; IPR009006; Racem_decarbox_C.
DR Pfam; PF00842; Ala_racemase_C; 1.
DR Pfam; PF01168; Ala_racemase_N; 1.
DR PRINTS; PR00992; ALARACEMASE.
DR TIGR; TIGR00492; alt; 1.
```

```
DR PROSITE; PS00395; ALANINE_RACEMASE; UNKNOWN 1.
KW Complete proteome; Isomerase; Pyridoxal phosphate.
SQ SEQUENCE 391 AA; 43371 MW; F8AAL73912483DF4 CRC64;
Query Match 0.8%; Score 8; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 988 ILGYTSVA 995
Db 86 ILGYTSVA 93
|||||
RESULT 53
Q6HJQ1 PRELIMINARY; PRT; 391 AA.
AC Q6HJQ1;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Alanine racemase (EC 5.1.1.1).
GN Name=dal; OrderedLocusNames=BT9727_1894;
OS Bacillus thuringiensis (subsp. konkukian).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=180856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=97-27;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RT "Complete genome sequence of Bacillus thuringiensis 97-27.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBSJ databases.
CC -!- CATALYTIC ACTIVITY: L-alanine = D-alanine.
CC -!- COPACTORY: Pyridoxal phosphate (By similarity).
CC -!- SIMILARITY: Belongs to the alanine racemase family.
DR EMBL; AE017355; AAY63484.1; -
DR GO; GO:0008784; F:alanine racemase activity; IEA.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR GO; GO:0030170; F:pyridoxal phosphate binding; IEA.
DR GO; GO:0006522; P:alanine metabolism; IEA.
DR InterPro; IPR000821; Ala_racemase.
DR InterPro; IPR011079; Ala_racemase_C.
DR InterPro; IPR001608; Ala_racemase_N.
DR InterPro; IPR009006; Racem_decarbox_C.
DR Pfam; PF00842; Ala_racemase_C; 1.
DR Pfam; PF01168; Ala_racemase_N; 1.
DR PRINTS; PR00992; ALARACEMASE.
DR TIGR; TIGR00492; alt; 1.
DR PROSITE; PS00395; ALANINE_RACEMASE; UNKNOWN 1.
KW Complete proteome; Isomerase; Pyridoxal phosphate.
SQ SEQUENCE 391 AA; 43419 MW; 2209861B4C733C21 CRC64;
Query Match 0.8%; Score 8; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 988 ILGYTSVA 995
Db 86 ILGYTSVA 93
|||||
RESULT 54
ACK2_RHIME STANDARD; PRT; 393 AA.
AC ACK2_RHIME; Q9X4G2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Acetate kinase (EC 2.7.2.1) (Acetokinas).
GN Name=ackA;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
```

OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=104A14;
 RX MEDLINE=99194732; PubMed=10094701;
 RA Summers M.L., Denton M.C., McDermott T.R.;
 RT "Genes coding for phosphotransacetylase and acetate kinase in
 RT Sinorhizobium meliloti are in an operon that is inducible by phosphate
 RT stress and controlled by phoB.";
 RJ J. Bacteriol. 181:2217-2224(1999).
 RN [2]
 RP SEQUENCE OF 207-314 FROM N.A.
 RC STRAIN=104A14;
 RX MEDLINE=9902212; PubMed=9805396;
 RA Summers M.L., Elkins J.G., Elliot B.A., McDermott T.R.;
 RT "Expression and regulation of phosphate stress inducible genes in
 RT Sinorhizobium meliloti.";
 RJ Mol. Plant Microbe Interact. 11:1094-1101(1998).
 CC -!- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.
 CC -!- PATHWAY: Conversion of acetate to acetyl-CoA; first step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the acetate kinase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL; AF095903; AAD24358.1; -;
 DR EMBL; AF074452; AAD42996.1; -;
 DR HSSP; P38502; 1G99.
 DR HAMAP; MF_00020; -; 1.
 DR InterPro; IPR000890; Acetate kin.
 DR InterPro; IPR004372; AcK.
 DR Pfam; PF00871; Acetate kinase; 1.
 DR PRINTS; PR00471; ACETATEKINASE.
 DR TIGRFAMs; TIGR00016; acK; 1.
 DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
 DR PROSITE; PS01076; ACETATE_KINASE_2; FALSE_NEG.
 KW Kinase; Transferase.
 FT CONFLICT 209 212 SGAS -> AEFR (in Ref. 2).
 FT CONFLICT 227 238 LHRPSGCRWDT -> GFTALDGLPMGTR (in Ref.
 FT SEQUENCE 393 AA; 42135 MW; 48FD185524CD6D4C CRC64;
 Query Match 0.8%; Score 8; DB 1; Length 393;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 169 AADGTOLI 176
 DB 47 AADGTOLI 54
 RESULT 55
 Q6BJ95 PRELIMINARY; PRT; 416 AA.
 AC Q6BJ95;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Debaryomyces hansenii chromosome G of strain CBS767 of Debaryomyces
 DE hansenii.
 GN ORFNames=DEHA0G04763g;
 OS Debaryomyces hansenii CBS767.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
 OX NCBI_TaxID=284592;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=CBST767;
 RG Genolevures;
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neveglise C., Talla E.,
 RA Goffard N., Frangeul L., Aigle M., Anchoard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
 RA Boissarie A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jaumiaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
 RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissbach J.,
 RA Winkler P., Souciet J.L.;
 RT "Genome evolution in yeasts.";
 RL Nature 430:35-44(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBST767;
 RA Genoscope;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBSJ databases.
 CC -!- SIMILARITY: Belongs to the AAA ATPase family.
 DR EMBL; CR382139; CAG90179.1; -;
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0030163; P:protein catabolism; IEA.
 DR InterPro; IPR005917; 26S p45.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003959; AAA_ATPase_cent.
 DR InterPro; IPR003960; AAA_sub.
 DR Pfam; PF00004; AAA; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR01242; 26Sp45; 1.
 DR PROSITE; PS00674; AAA; 1.
 KW ATP-binding.
 SQ SEQUENCE 416 AA; 46706 MW; 9C4C95550B5F33DE CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 416;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 795 LEKELDLL 802
 DB 42 LEKELDLL 49
 RESULT 56
 AROA_STRMU STANDARD; PRT; 427 AA.
 AC Q8DUV8;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
 GN Name=aroA; OrderedLocusNames=SMU.784;
 OS Streptococcus mutans;
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UA159 / ATCC 700610 / Serotype c;
 RX MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,

```

RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RL pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE014920; AAN58503.1; -.
DR HAMAP; MF_00210; -.
DR InterPro; IPR006264; AroA.
DR Pfam; PF00275; EPSP_synthase.
DR ProDom; PD001867; EPSP_synth; 1.
DR TIGRFAMs; TIGR01356; aroA; 1.
DR PROSITE; PS00104; EPSP_SYNTHASE 1; 1.
DR PROSITE; PS00885; EPSP_SYNTHASE 2; 1.
KW Aromatic amino acid biosynthesis; Complete proteome; Transferase.
SQ SEQUENCE 427 AA; 46088 MW; BF15D9C640B40AFC CRC64;

Query Match 0.8%; Score 8; DB 1; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LDVQKEI 214
|||
DB 207 LDVQKEI 214

RESULT 57
Q8CCKS PRELIMINARY; PRT; 427 AA.
AC Q8CCKS;
DT 01-MAR-2003 [TrEMBLrel. 23, Created]
DT 01-MAR-2003 [TrEMBLrel. 23, Last sequence update]
DT 01-OCT-2003 [TrEMBLrel. 25, Last annotation update]
DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
DE enriched library, clone:6430703L12 product:hypothetical protein, full
DE insert sequence.
GN Name=Rtnn;
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=21085650; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;

```

```

RA The FANTOM Consortium;
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 70,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai C., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tomaru A., Toyata T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK032605; BAC27946.1; -.
DR MGI; MGI:2179288; Rtnn.
DR GO; GO:0007368; P:determination of left/right symmetry; IMP.
KW Hypothetical protein.
SQ SEQUENCE 427 AA; 47304 MW; DCF5DB1C96026C9D CRC64;

Query Match 0.8%; Score 8; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 824 LKTPPLP 831
|||
DB 168 LKTPPLP 175

RESULT 58
Q97KAI PRELIMINARY; PRT; 429 AA.
AC Q97KAI;
DT 01-OCT-2001 [TrEMBLrel. 18, Created]
DT 01-OCT-2001 [TrEMBLrel. 18, Last sequence update]
DE Predicted membrane protein.
GN OrderedLocustName=CAC1018;
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21399325; PubMed=11466286;
 RX DOI=10.1128/JB.183.16.4823-4838.2001;
 RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatunov R.L., Sabache F., Doucette-Stamm L.A., Soucaille P.,
 RA Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum";
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL; AE007617; AAK78994.1; -;
 DR PIR; G97025; G97025.
 DR InterPro; IPR011041; Quino_gluc_DH.
 KW Complete proteome.
 SQ SEQUENCE 429 AA; 48545 MW; 114D12A7262514AA CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 841 FDKSGKLI 848
 |||||
 Db 239 FDKSGKLI 246
 |||||
 RESULT 59
 Q9NKK1 PRELIMINARY; PRT; 441 AA.
 AC Q9NKK1
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Peroxisomal multifunctional enzyme MFE homolog.
 GN Name=mfeA;
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KAX4;
 RX MEDLINE=22680914; PubMed=12796309; DOI=10.1128/EC.2.3.638-645.2003;
 RA Matsuoka S., Saito T., Kuwayama H., Morita N., Ochiai H., Maeda M.;
 RT "MFE1, a member of the peroxisomal hydroxyacyl coenzyme A
 RT dehydrogenase family, affects fatty acid metabolism necessary for
 RT morphogenesis in Dictyostelium spp.";
 RL Eukaryotic Cell 2:638-645(2003).
 CC -!- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases
 CC (SDR) family.
 DR EMBL; AB042104; BAA94961.1; -;
 DR HSSP; P97852; 1GZ6.
 DR DictyBase; DDB0201628; mfeA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0003498; F:sterol carrier activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR002198; ADH short.
 DR InterPro; IPR002347; Adh_short_C2.
 DR InterPro; IPR003033; SCP2.
 DR Pfam; PF00106; adh_short; 1.
 DR Pfam; PF02036; SCP2; 1.
 DR PRINTS; PR00081; GDHRDH.
 DR PRINTS; PR00080; SDRFAMILY.
 KW Oxidoreductase.
 SQ SEQUENCE 441 AA; 46387 MW; 2CEF3D9779D93BCA CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 441;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 79 TPEIKDN 86
 |||||
 Db 259 TPEIKDN 266
 |||||

RESULT 60
 ID ARLY_BACHD STANDARD; PRT; 458 AA.
 AC Q9K821;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Argininosuccinate lyase (EC 4.3.2.1) (Argininosuccinase) (ASAL).
 GN Name=argH; OrderedLocNames=BH3186;
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 RX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX STRAIN=20512582; PubMed=11058132; DOI=10.1093/nar/28.21.4317;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 CC -!- CATALYTIC ACTIVITY: (N(omega)-L-arginino)succinate = fumarate + L-
 CC arginine.
 CC -!- PATHWAY: Arginine biosynthesis; eighth (last) step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -!- SIMILARITY: Belongs to the lyase 1 family. Argininosuccinate lyase
 CC subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; AP001518; BAB06905.1; -;
 DR PIR; B84048; B84048.
 DR HSSP; P24058; 1DCN.
 DR HAMAP; MF_00006; -; 1.
 DR InterPro; IPR009049; argH.
 DR InterPro; IPR003031; D_crySTALLIN.
 DR InterPro; IPR000362; Fumarate lyase.
 DR InterPro; IPR008948; L-Aspartase-like.
 DR Pfam; PF02006; Lyase_1; 1.
 DR PRINTS; PR00145; DCRYSTALLIN.
 DR PRINTS; PR00149; FUMRATELYASE.
 DR TIGRFAMs; TIGR00838; argH; 1.
 DR PROSITE; PS00163; FUMARATE LYASES; 1.
 KW Arginine biosynthesis; Complete proteome; Lyase.
 SQ SEQUENCE 458 AA; 51104 MW; CFB32B39DF4FC40 CRC64;
 Query Match 0.8%; Score 8; DB 1; Length 458;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 750 KAIKAEK 757
 |||||
 Db 447 KAIKAEK 454
 |||||
 RESULT 61
 Q7VVB6 PRELIMINARY; PRT; 463 AA.
 ID Q7VVB6
 AC Q7VVB6
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Exodeoxyribonuclease large subunit (EC 3.1.11.6).
 GN Name=xseA; OrderedLocNames=Bp2762;
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=520;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitch E., Rutter S., Sanders M., Saunders S., Stevens K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares D., Seeger K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
 RA "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640419; CAE3037.1; --
 DR GO; GO:0009318; C:exodeoxyribonuclease VII complex; IEA.
 DR GO; GO:0008855; F:exodeoxyribonuclease VII activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0003676; P:nucleic acid binding; IEA.
 DR GO; GO:0006308; P:DNA catabolism; IEA.
 DR GO; GO:0006308; P:DNA catabolism; IEA.
 DR InterPro; IPR003753; Exonuc VII L.
 DR InterPro; IPR008994; Nucleic acid OB.
 DR Pfam; PF01336; tRNA_anti. 1.
 DR TIGRFAMs; TIGR00237; xseA; 1.
 DR Complete proteome; Hydrolase.
 KW Complete proteome; Hydrolase.
 SQ SEQUENCE 463 AA; 5024 MW; AAFB2CCEB494AD79 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 768 QOLLERSI 775
 DB 27 QOLLERSI 34
 RESULT 62
 ID Q7W7F4 PRELIMINARY; PRT; 463 AA.
 AC Q7W7F4
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Exodeoxyribonuclease large subunit (EC 3.1.11.6).
 GN Name=xseA; OrderedLocNames=BPP2566;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitch E., Rutter S., Sanders M., Saunders S., Stevens K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares D., Seeger K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
 RA "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640443; CAE32508.1; --
 DR GO; GO:0009318; C:exodeoxyribonuclease VII complex; IEA.
 DR GO; GO:0008855; F:exodeoxyribonuclease VII activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0003676; P:nucleic acid binding; IEA.
 DR GO; GO:0006308; P:DNA catabolism; IEA.
 DR InterPro; IPR003753; Exonuc VII L.
 DR InterPro; IPR008994; Nucleic acid OB.
 DR Pfam; PF01336; tRNA_anti. 1.
 DR TIGRFAMs; TIGR00237; xseA; 1.
 DR Complete proteome; Hydrolase.
 KW Complete proteome; Hydrolase.
 SQ SEQUENCE 463 AA; 50263 MW; 4CC6962BFD491BA4 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 768 QOLLERSI 775
 DB 27 QOLLERSI 34
 RESULT 62
 ID Q7W7F4 PRELIMINARY; PRT; 463 AA.
 AC Q7W7F4
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Exodeoxyribonuclease large subunit (EC 3.1.11.6).
 GN Name=xseA; OrderedLocNames=BPP2566;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=12822 / ATCC BAA-587;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitch E., Rutter S., Sanders M., Saunders S., Stevens K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares D., Seeger K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
 RA "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640430; CAE37860.1; --

DR GO; GO:0009318; C:exodeoxyribonuclease VII complex; IEA.
 DR GO; GO:0008855; F:exodeoxyribonuclease VII activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0003676; P:nucleic acid binding; IEA.
 DR GO; GO:0006308; P:DNA catabolism; IEA.
 DR InterPro; IPR003753; Exonuc VII L.
 DR InterPro; IPR008994; Nucleic acid OB.
 DR Pfam; PF01336; tRNA_anti. 1.
 DR TIGRFAMs; TIGR00237; xseA; 1.
 DR Complete proteome; Hydrolase.
 KW Complete proteome; Hydrolase.
 SQ SEQUENCE 463 AA; 50263 MW; 4CC6962BFD491BA4 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 768 QOLLERSI 775
 DB 27 QOLLERSI 34
 RESULT 63
 ID Q7WKU2 PRELIMINARY; PRT; 463 AA.
 AC Q7WKU2
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Exodeoxyribonuclease large subunit (EC 3.1.11.6).
 GN Name=xseA; OrderedLocNames=BB2011;
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitch E., Rutter S., Sanders M., Saunders S., Stevens K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares D., Seeger K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
 RA "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640443; CAE32508.1; --
 DR GO; GO:0009318; C:exodeoxyribonuclease VII complex; IEA.
 DR GO; GO:0008855; F:exodeoxyribonuclease VII activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0003676; P:nucleic acid binding; IEA.
 DR GO; GO:0006308; P:DNA catabolism; IEA.
 DR InterPro; IPR003753; Exonuc VII L.
 DR InterPro; IPR008994; Nucleic acid OB.
 DR Pfam; PF01336; tRNA_anti. 1.
 DR TIGRFAMs; TIGR00237; xseA; 1.
 DR Complete proteome; Hydrolase.
 KW Complete proteome; Hydrolase.
 SQ SEQUENCE 463 AA; 50263 MW; 4CC6962BFD491BA4 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 768 QOLLERSI 775
 DB 27 QOLLERSI 34


```

RESULT 64
Q9YE79          PRELIMINARY;          PRT;    478 AA.
AC Q9YE79;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE0691.
GN OrderedLocuNames=AP0691;
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79667.1; -.
DR PIR; C72658; C72658.
DR InterPro; IPR001992; Bact_sec_systII.
DR Pfam; PF00482; GSP11.F; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 478 AA; 49725 MW; F05565A05C2A3F58 CRC64;

Query Match          0.8%; Score 8; DB 2; Length 478;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 LKVIGEAS 333
DB 406 LKVIGEAS 413

RESULT 65
Q9NW94          PRELIMINARY;          PRT;    483 AA.
AC Q9NW94;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ10206.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryo;
RX PubMed=14702039; DOI=10.1038/ngl1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiori T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

```

```

RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyaama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamagaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs";
RL Nat. Genet. 36:40-45(2004).
RL EMBL; AK001068; BAA91489.1; -.
SQ SEQUENCE 483 AA; 54325 MW; 86CABC2ED5C76176 CRC64;

Query Match          0.8%; Score 8; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 945 EKSSSES 952
DB 444 EKSSSES 451

RESULT 66
Q6LUN0          PRELIMINARY;          PRT;    485 AA.
AC Q6LUN0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative sigma-54 dependent transcriptional regulator.
GN Name=CPSP; OrderedLocuNames=PBPA0572;
OS Photobacterium profundum (Photobacterium sp. (strain SS9)).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Photobacterium.
OX NCBI_TaxID=74109;
RN [1]
RP SEQUENCE FROM N.A.
RA Vezzi A., Campanaro S., D'Angelo M., Simonato F., Vitulo N., Lauro F.,
RA Cestaro A., Malacrida G., Simonati B., Cannata N., Bartlett D.,
RA Valle G.;
RT "Genome analysis of Photobacterium profundum reveals the complexity of
RT high pressure adaptations.";
RL Submitted (MAR-2004) to the EMBL/GenBank/DBSJ databases.
CC -1- SIMILARITY: Contains 1 sigma-54 factor interaction ATP-binding
CC domain.
DR EMBL; CR378664; CAG18995.1; -.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0000370; F:transcription factor activity; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .); IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR011006; CheY-like.
DR InterPro; IPR008931; FIS-like.
DR InterPro; IPR002197; HTH_Fis.
DR InterPro; IPR002078; Sig54_interact.
DR Pfam; PF02954; HTH_8; 1.
DR Pfam; PF00158; Sigma54_activat; 1.
DR PRINTS; PR01590; HTHFIS.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR01199; HTH_fis; 1.
DR PROSITE; PS00045; SIGMA54_INTERACT 4; 1.
KW ATP-binding; Complete proteome; DNA-binding; Transcription;
KW Transcription regulation.
SQ SEQUENCE 485 AA; 54241 MW; 18D39CEBD9DBE23C CRC64;

Query Match          0.8%; Score 8; DB 2; Length 485;

```

```

Best Local Similarity 100.0%; Pred. No. 2e+02; 0; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy 613 SIKTIREE 620
Db 427 SIKTIREE 434

RESULT 67
Q42989
ID Q42989 PRELIMINARY; PRT; 488 AA.
AC Q42989;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Beta-amylose (EC 3.2.1.2).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RC STRAIN=Tainung 67;
RP SEQUENCE FROM N.A.
RA Chen J.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; L10345; AAA33898.1; -.
DR HSSP; F16098; 1BIY.
DR Gramene; Q42989; -.
DR GO; GO:0016161; F:beta-amylose activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETA_AMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
DR PROSITE; PS00136; SUBTILASE ASP; UNKNOWN 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 488 AA; 55093 MW; 30350B1AE7A8DADF CRC64;

Query Match 0.8%; Score 8; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 68
Q42990
ID Q42990 PRELIMINARY; PRT; 488 AA.
AC Q42990;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Beta-amylose (EC 3.2.1.2).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RC STRAIN=Tainung 67;
RP SEQUENCE FROM N.A.
RA Chen J.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; L10345; AAA33898.1; -.
DR HSSP; F16098; 1BIY.
DR Gramene; Q42989; -.
DR GO; GO:0016161; F:beta-amylose activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETA_AMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
DR PROSITE; PS00136; SUBTILASE ASP; UNKNOWN 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 488 AA; 55093 MW; 30350B1AE7A8DADF CRC64;

Query Match 0.8%; Score 8; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 69
Q42991
ID Q42991 PRELIMINARY; PRT; 488 AA.
AC Q42991;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Putative beta-amylose.
GN Name=OJ1729_E01.25;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RC STRAIN=OJ1729_E01.25;
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT clone:OJ1729_E01.25";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AP005156; BAC83773.1; -.
DR HSSP; F16098; 1BIY.
DR GO; GO:0016161; F:beta-amylose activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETA_AMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.

```

```
DR PROSITE; P500679; BETA_AMYLASE_2; 1.
DR PROSITE; P500136; SUBTILASE ASP; UNKNOWN 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 488 AA; 55117 MW; 9E3F42A8F93EC5D6 CRC64;

Query Match      0.8%; Score 8; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 70
Q8WPP2 PRELIMINARY; PRT; 490 AA.
AC Q8WPP2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 6-phospho-1-fructokinase (EC 2.7.1.11).
GN Name=pfk;
OS Trypanoplasma borreli.
OC Eukaryota; Euklenozoa; Kinetoplastida; Bodonidae; Trypanoplasma.
OX NCBI_TaxID=5710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22168988; PubMed=12180974;
RA Lopez C., Chevallier N., Hannaert V., Rigden D.J., Michels P.A.,
RA Ramirez J.L.;
RT "Leishmania donovani phosphofructokinase: gene characterization,
RT biochemical properties and structure-modelling studies.";
RL Eur. J. Biochem. 269:3978-3989(2002).
DR EMBL; AJ310928; CAC84571.1; -.
DR GO; GO:0005945; C:6-phosphofructokinase complex; IEA.
DR GO; GO:0003872; F:6-phosphofructokinase activity; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008096; P:glycolysis; IEA.
DR Pfam; PF00365; PFK; 1.
DR PRINTS; PR00476; PHFRCTKINASE.
KW Kinase; Transferase.
SQ SEQUENCE 490 AA; 53376 MW; 62360A4071A700DF CRC64;

Query Match      0.8%; Score 8; DB 2; Length 490;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 632 VASALSHI 639
Db 287 VASALSHI 294

RESULT 71
Q6C9U7 PRELIMINARY; PRT; 492 AA.
AC Q6C9U7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to w|NCU01449.1 Neurospora crassa NCU01449. 1 hypothetical
DE protein.
ORFNames=YAI0D082509;
OS Yarrowia lipolytica Cl1999.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=284591;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL1B99;
RG Genolevures;

RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boissane A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CL1B99;
RC Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; CR382130; CAG80753.1; -.
DR GO; GO:0003935; F:GTP cyclohydrolase II activity; IEA.
DR GO; GO:0009231; P:vitamin B2 biosynthesis; IEA.
KW Hypothetical protein.
SQ SEQUENCE 492 AA; 54208 MW; EB4A540581E29863 CRC64;

Query Match      0.8%; Score 8; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 879 YHALAVAT 886
Db 134 YHALAVAT 141

RESULT 72
AMVB WHEAT
ID AMVB WHEAT STANDARD; PRT; 503 AA.
AC P93594;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Beta-amylase (EC 3.2.1.2) (1,4-alpha-D-glucan maltohydrolase).
GN Name=BMV1; Synonyms=AMV1;
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Star; TISSUE=Leaf;
RC Wagner G.B., Haeger K.-P., Ziegler P.;
RT "Nucleotide sequence of a cDNA from wheat leaves encoding ubiquitous
RT beta-amylase.";
RL (er) Plant Gene Register PGR96-123.
CC -| CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -| SIMILARITY: Belongs to the glycosyl hydrolase 14 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X98504; CAA67128.1; -.
CC HSP; P16098; 1BIY.
DR InterPro; IPR001554; Glyco_hydro_14.
```

```

DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETA_AMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE 1; 1.
DR PROSITE; PS00679; BETA_AMYLASE 2; 1.
KW Glycosidase; Hydrolase; Polysaccharide degradation.
FT ACT_SITE 184 184 By similarity.
FT ACT_SITE 378 378 By similarity.
SQ SEQUENCE 503 AA; 56611 MW; 7189E1533A6C0F73 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 73
GPMI_MYCGA STANDARD; PRT; 503 AA.
AC Q7NAQ5;
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE 2,3-bisphosphoglycerate-independent phosphoglycerate mutase
DE (EC 5.4.2.1) (Phosphoglyceromutase) (BPG-independent PGAM) (iPGM).
GN Name=gpmi; OrderedLocNames=MYCGA5800; ORFNames=MGA_0356;
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K (low);
RX MEDLINE=2830409; PubMed=12949158; DOI=10.1099/mic.0.26427-0;
RA Papazizi L., Gorton T.S., Kutish G., Markham P.F., Browning G.F.,
RA Nguyen D.K., Swartzell S., Madan A., Mahairas G., Geary S.J.;
RT "The complete genome sequence of the avian pathogen Mycoplasma
gallisepticum strain R (low).";
RL Microbiology 149:2307-2316(2003).
CC -1- FUNCTION: Catalyzes the interconversion of 2-phosphoglycerate and
CC 3-phosphoglycerate (By similarity).
CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate = 3-phospho-D-glycerate.
CC -1- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Glycolysis.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SIMILARITY: Belongs to the BPG-independent phosphoglycerate mutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016969; AAP56930.1; -.
CC HAMAP; MF_01038; -.
CC InterPro; IPR004456; APGAM.
CC InterPro; IPR011258; iPGM N.
CC InterPro; IPR006124; Metalloenzyme.
CC InterPro; IPR005995; Pgm_bpd_ind.
CC Pfam; PF06415; iPGM N; 1.
CC Pfam; PF01676; Metalloenzyme; 1.
CC PIRSF; PIRSF001492; iPGAM; 1.
CC ProDom; PD004429; Pgm_bpd_ind; 1.
CC TIGRfam; TIGR01307; pgm_bpd_ind; 1.
KW Complete proteome; Glycolysis; Isomerase; Manganese; Metal-binding.
FT ACT_SITE 60 60 Phosphoserine intermediate (By
FT similarity).
FT METAL 10 10 Manganese 2 (By similarity).

```

```

FT METAL 60 60 Manganese 2 (By similarity).
FT METAL 396 396 Manganese 1 (By similarity).
FT METAL 400 400 Manganese 1 (By similarity).
FT METAL 437 437 Manganese 2 (By similarity).
FT METAL 438 438 Manganese 2 (By similarity).
FT METAL 455 455 Manganese 1 (By similarity).
SQ SEQUENCE 503 AA; 56579 MW; 3FDD2C55BB396979 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 532 HINHIYAI 539
Db 123 HINHIYAI 130

RESULT 74
Q08335 PRELIMINARY; PRT; 503 AA.
ID Q08335;
AC Q08335;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Beta-amylase.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94151427; PubMed=8108499; DOI=10.1104/pp.102.1.315;
RA Sadoweki J., Rorat T., Cooke R., Deleensy M.;
RT "Nucleotide sequence of a cDNA clone encoding ubiquitous beta-amylase
in rye (Secale cereale L.).";
RL Plant Physiol. 102:315-316(1993).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -1- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
CC
CC EMBL; Z11772; CAA77817.1; -.
CC PIR; JQ2248; JQ2248.
CC HSP; P16098; IBIY.
CC DR GO; GO:0016161; F:beta-amylase activity; IEA.
CC DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
CC DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
CC DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
CC DR InterPro; IPR001554; Glyco_hydro_14.
CC DR InterPro; IPR001371; Glyco_hydro_14B.
CC Pfam; PF01373; Glyco_hydro_14; 1.
CC PRINTS; PR00750; BETA_AMYLASE.
CC DR PROSITE; PR00842; GLHYDLASE14B.
CC DR PROSITE; PS00506; BETA_AMYLASE 1; 1.
CC DR PROSITE; PS00679; BETA_AMYLASE 2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 503 AA; 56759 MW; D788B79A9B8D9523 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 75
Q6SNP7 PRELIMINARY; PRT; 517 AA.
ID Q6SNP7;
AC Q6SNP7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)

```

DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE Beta-amylase 1 (Fragment).
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Clark S.E., Hayes P.M., Henson C.A.;
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
 CC polysaccharides so as to remove successive maltose units from the
 CC non-reducing ends of the chains.
 CC -|- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
 DR EMBL; AY454398; RAR18251.1; -.
 DR HSSP; P16098; 1BY.
 DR GO; GO:0016161; F:beta-amylase activity; IEA.
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:000272; P:polysaccharide catabolism; IEA.
 DR InterPro; IPR001554; Glyco_hydro_14.
 DR InterPro; IPR001371; Glyco_hydro_14B.
 DR Pfam; PF01373; Glyco_hydro_14; 1.
 DR PRINTS; PR00750; BETAMYLASE.
 DR PRINTS; PR00842; GLHYDLASE14B.
 DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
 DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
 DR Carbohydrate metabolism; Glycosidase; Hydrolase;
 KW Polysaccharide degradation.
 FT NON_TER 1
 FT NON_TER 517 517
 SQ SEQUENCE 517 AA; 57635 MW; F41B4950559871BB CRC64;

 Query Match 0.8%; Score 8; DB 2; Length 517;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 804 GLVEGKGP 811
 Db |||||
 43 GLVEGKGP 50

 RESULT 76
 Q7MD66 PRELIMINARY; PRT; 528 AA.
 ID Q7MD66; PRELIMINARY; PRT; 528 AA.
 AC Q7MD66;
 DT 01-MAR-2004 (TRENBLrel. 26, Created)
 DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Putative N-acyl-D-glutamate deacylase protein.
 GN OrderedLocusNames=VVAL170;
 OS Vibrio vulnificus (strain VJ016).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=196600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=1456965; DOI=10.1101/gr.1295503;
 RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
 RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
 RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-P.;
 RT "Comparative genome analysis of Vibrio vulnificus, a marine
 RT pathogen.";
 RL Genome Res. 13:2577-2587(2003).
 DR EMBL; AP005349; BAC97196.1; -.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR006680; Amidohydro_1
 DR InterPro; IPR011059; Metallo_hydrolase.
 DR Pfam; PF01979; Amidohydro_1; 1.
 KW Complete proteome.
 SQ SEQUENCE 528 AA; 58175 MW; 584BF68E20C68097 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 528;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 689 EKILKLIIE 696
 Db |||||
 186 EKILKLIIE 193

 RESULT 77
 Q9FUK6 PRELIMINARY; PRT; 533 AA.
 ID Q9FUK6;
 AC Q9FUK6;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Beta-amylase (Fragment).
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ma Y., Evans E.D., Logue S.J., Langridge P.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
 CC polysaccharides so as to remove successive maltose units from the
 CC non-reducing ends of the chains.
 CC -|- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
 DR EMBL; AF300800; AAG25638.1; -.
 DR HSSP; P16098; 1BY.
 DR GO; GO:0016161; F:beta-amylase activity; IEA.
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:000272; P:polysaccharide catabolism; IEA.
 DR InterPro; IPR001554; Glyco_hydro_14.
 DR InterPro; IPR001371; Glyco_hydro_14B.
 DR Pfam; PF01373; Glyco_hydro_14; 1.
 DR PRINTS; PR00750; BETAMYLASE.
 DR PRINTS; PR00842; GLHYDLASE14B.
 DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
 DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
 KW Carbohydrate metabolism; Glycosidase; Hydrolase;
 KW Polysaccharide degradation.
 FT NON_TER 1
 FT NON_TER 533 533
 SQ SEQUENCE 533 AA; 59312 MW; C0B9780B0F13F48C CRC64;

 Query Match 0.8%; Score 8; DB 2; Length 533;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 804 GLVEGKGP 811
 Db |||||
 53 GLVEGKGP 60

 RESULT 78
 Q9FUK7 PRELIMINARY; PRT; 533 AA.
 ID Q9FUK7;
 AC Q9FUK7;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Beta-amylase (Fragment).
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.

```
RA Ma Y., Evans E.D., Logue S.J., Langridge P.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBSJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AF300799; AAG25637.1; -.
DR HSSP; P16098; 1BY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001554; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE 1; 1.
DR PROSITE; PS00679; BETA_AMYLASE 2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
FT NON_TER 1
SQ SEQUENCE 533 AA; 59387 MW; 793031ED53FFACB5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 533;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
|||||
DB 53 GLVEGKGP 60

RESULT 79
AMYB HORVU STANDARD; PRT; 535 AA.
ID AMYB HORVU STANDARD; PRT; 535 AA.
AC P16098;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Beta-amylase (EC 3.2.1.2) (1,4-alpha-D-glucan maltohydrolase).
GN Name=BMV1; Synonyms=AMVB;
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Endosperm;
RX MEDLINE=88082785; PubMed=2446870;
RA Kreis M., Williamson M., Buxton B., Pywell J., Hejgaard J.,
RA Svendsen I.;
RT "Primary structure and differential expression of beta-amylase in
RT normal and mutant barleys";
RL Eur. J. Biochem. 169:517-525 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Haruna Two-rows;
RX MEDLINE=96068953; PubMed=8534999;
RA Yoshigi N., Okada Y., Sahara H., Tamaki T.;
RT "A structural gene encoding beta-amylase of barley";
RL Biosci. Biotechnol. Biochem. 59:1991-1993 (1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Haruna Two-rows;
RX MEDLINE=94245663; PubMed=8188635;
RA Yoshigi N., Okada Y., Sahara H., Koshino S.;
RT "PCR cloning and sequencing of the beta-amylase cDNA from barley.";
RL J. Biochem. 115:47-51 (1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 5-504.
RX MEDLINE=99141375; PubMed=9918723; DOI=10.1006/jmbi.1998.2379;
```

```
RA Mikami B., Yoon H.-J., Yoshigi N.;
RT "The crystal structure of the sevenfold mutant of barley beta-amylase
RT with increased thermostability at 2.5 A resolution.";
RL J. Mol. Biol. 285:1235-1243 (1999).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 14 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; X52321; CAA36556.1; -.
DR EMBL; D49999; BAA08741.1; -.
DR EMBL; D21349; BAA04815.1; -.
DR PIR; S00222; S00222.
DR PDB; 1BY; X-ray; A=5-504.
DR InterPro; IPR001554; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE 1; 1.
DR PROSITE; PS00679; BETA_AMYLASE 2; 1.
KW 3D-structure; Direct protein sequencing; Glycosidase; Hydrolase;
KW Polysaccharide degradation; Repeat.
FT ACT_SITE 184 By similarity.
FT ACT_SITE 378 By similarity.
FT DOMAIN 489 532 4 X 11 AA tandem repeats.
FT REPEAT 489 499 1.
FT REPEAT 500 510 2.
FT REPEAT 511 521 3.
FT REPEAT 522 532 4 (approximate).
FT CONFLICT 233 233 V -> A (in Ref. 2 and 3).
FT CONFLICT 347 347 L -> S (in Ref. 2 and 3).
FT CONFLICT 527 527 I -> M (in Ref. 2 and 3).
FT HELIX 6 8
FT STRAND 11 15
FT TURN 18 19
FT TURN 23 24
FT TURN 29 29
FT HELIX 30 42
FT TURN 43 44
FT STRAND 47 53
FT TURN 54 56
FT HELIX 57 60
FT TURN 62 63
FT TURN 68 80
FT TURN 81 81
FT STRAND 83 89
FT STRAND 93 93
FT TURN 97 98
FT STRAND 103 103
FT HELIX 107 115
FT HELIX 117 119
FT STRAND 120 122
FT TURN 124 125
FT STRAND 128 133
FT HELIX 135 137
FT TURN 138 139
FT TURN 142 144
FT HELIX 148 169
FT TURN 170 170
FT STRAND 172 177
FT HELIX 181 183
FT STRAND 184 184
FT HELIX 192 194
FT TURN 195 195
```

TURN 198 199
 FT TURN 209 221
 FT TURN 222 223
 FT TURN 225 226
 FT TURN 231 232
 FT TURN 236 237
 FT TURN 240 242
 FT TURN 244 246
 FT TURN 248 249
 FT TURN 251 253
 FT TURN 255 283
 FT TURN 284 285
 FT STRAND 289 294
 FT TURN 299 300
 FT TURN 307 312
 FT TURN 313 313
 FT TURN 322 323
 FT TURN 324 330
 FT TURN 331 334
 FT STRAND 336 339
 FT TURN 342 343
 FT TURN 346 348
 FT TURN 351 354
 FT TURN 357 370
 FT TURN 371 372
 FT STRAND 375 377
 FT TURN 386 396
 FT TURN 398 399
 FT TURN 403 404
 FT STRAND 413 416
 FT TURN 421 423
 FT TURN 426 439
 FT TURN 440 442
 FT TURN 464 467
 FT TURN 468 470
 SQ SEQUENCE 535 AA; 59647 MW; FFDA4CED53E9A89E CRC64;

Query Match 0.8%; Score 8; DB 1; Length 535;
 Best Local Similarity 100.0%; Pred.No.2.2e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
 |||||
 Db 55 GLVEGKGP 62

RESULT 80
 ID P82993 PRELIMINARY; PRT; 535 AA.
 AC P82993;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Beta-amylase precursor (EC 3.2.1.2) (Beta-Amyl) (1,4-alpha-D-glucan
 maltohydrolase).
 GN Name=BMV1; Synonyms=AMYB;
 OS Hordeum spontaneum (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=77009;
 RN [1]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RC STRAIN=NPGS P1296897; TISSUE=Endosperm;
 RX MEDLINE=98289114; PubMed=9625721; DOI=10.1104/pp.117.2.679;
 RA Erkkila M.J., Leah R., Ahokas H., Cameron-Mills V.;
 RT "Allele-dependent barley grain beta-amylase activity.";
 RL Plant Physiol. 117:679-685(1998).
 RN [2]
 RP PARTIAL SEQUENCE, FUNCTION, TISSUE SPECIFICITY, ACETYLATION, AND
 RP VARIANTS ASP-165; SER-254 AND GLN-472.
 RC STRAIN=cv. CP177146-33; TISSUE=Seed;
 RA Eglinton J.K., Lahnstein J., Shirley N., Evans D.E.;

RT "The structural basis for functional differences between allelic forms
 of barley (Hordeum vulgare) beta-amylase.";
 RL Eur. J. Biochem. 0:0-0(2001).
 CC -|- FUNCTION: Catalyzes the liberation of maltose from 1,4-alpha-D
 glucans.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
 polysaccharides so as to remove successive maltose units from the
 non-reducing ends of the chains.
 CC -|- SUBUNIT: Monomer (By similarity).
 CC -|- TISSUE SPECIFICITY: Endosperm.
 CC -|- POLYMORPHISM: There are at least three alleles; SD2H; SD1 and
 SD2L. The sequence of SD2H is shown here.
 CC -|- MISCELLANEOUS: The three alleles show different thermostabilities
 and different affinities for soluble starch.
 CC -|- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
 DR EMBL; AF061204; AAC67246.1; -.
 DR HSSP; P16098; IB1Y.
 DR GO; GO:0016161; F:beta-amylase activity; IEA.
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
 DR InterPro; IPR001554; Glyco_hydro_14.
 DR InterPro; IPR001371; Glyco_hydro_14; 1.
 DR Pfam; PF01373; Glyco_hydro_14; 1.
 DR PRINTS; PR00750; BETAAMYLASE.
 DR PRINTS; PR00842; GLHYDLASE14B.
 DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
 DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
 DR Acetylation; Glycosidase; Hydrolase; Polymorphism;
 KW Polysaccharide degradation; Repeat.
 FT PROPEP 1 2
 FT CHAIN 3 489 Removed in mature form.
 FT PROPEP 3 489 Beta-amylase.
 FT PROPEP 490 535 Removed in mature form.
 FT MOD_RES 3 3 N-acetylvaline.
 FT ACT_SITE 99 99 By similarity.
 FT ACT_SITE 184 184 By similarity.
 FT DOMAIN 489 532 4 X 11 AA tandem repeats.
 FT REPEAT 489 499 1.
 FT REPEAT 500 510 2.
 FT REPEAT 511 521 3.
 FT REPEAT 522 532 4 (Approximate).
 FT VARIANT 165 165 E -> D.
 FT VARIANT 254 254 T -> S.
 FT VARIANT 472 472 K -> Q.
 SQ SEQUENCE 535 AA; 59639 MW; 0B265EAEF061C9B CRC64;

Query Match 0.8%; Score 8; DB 2; Length 535;
 Best Local Similarity 100.0%; Pred.No.2.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
 |||||
 Db 55 GLVEGKGP 62

RESULT 81
 ID Q84T19 PRELIMINARY; PRT; 535 AA.
 AC Q84T19;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Endosperm-specific beta-amylase 1.
 GN Name=Bamy1;
 OS Hordeum vulgare var. distichum (Two-rowed barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=112509;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Clark S.E., Hayes P.M., Henson C.A.;
 RT "Effects of single nucleotide polymorphisms in beta-amylase alleles

```

RT from barley on functional properties of the enzymes." ;
RL Plant Physiol. Biochem. 41:798-804(2003)
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AF414082; AAO67356.1; -.
DR HSSP; P16098; 1BIY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001354; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 535 AA; 59639 MW; 338FDCFFCB66A51C CRC64;

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62

RESULT 82
Q84T20 ID Q84T20 PRELIMINARY; PRT; 535 AA.
AC Q84T20;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-perm-specific beta-amylase 1.
DE Name=Bmy1;
OS Hordeum vulgare var. distichum (Two-rowed barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=112509;
RN [1]
RP SEQUENCE FROM N.A.
RA Clark S.E., Hayes P.M., Henson C.A.;
RA "Effects of single nucleotide polymorphisms in beta-amylase alleles
RT from barley on functional properties of the enzymes." ;
RL Plant Physiol. Biochem. 41:798-804(2003)
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AF414081; AAO67355.1; -.
DR HSSP; P16098; 1BIY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001354; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 535 AA; 59572 MW; 4653050B0F05F0A7 CRC64;

from barley on functional properties of the enzymes." ;
RL Plant Physiol. Biochem. 41:798-804(2003)
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AF414082; AAO67356.1; -.
DR HSSP; P16098; 1BIY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001354; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 535 AA; 59639 MW; 338FDCFFCB66A51C CRC64;

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62

RESULT 83
Q9AVJ8 ID Q9AVJ8 PRELIMINARY; PRT; 535 AA.
AC Q9AVJ8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sdi beta-amylase.
DE Name=Bmy1;
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RA Li C., Langridge P., Zhang X., Eckstein P.E., Rosenagel B.G.,
RA Lance R.C., Lefol E.B., Lu M., Harvey B.L., Scoles G.J.;
RA Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RL -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
RL polysaccharides so as to remove successive maltose units from the
RL non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AB048949; BAB39391.1; -.
DR HSSP; P16098; 1BIY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001354; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 535 AA; 59572 MW; 4653050B0F05F0A7 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62

RESULT 84
Q9FSI3 ID Q9FSI3 PRELIMINARY; PRT; 535 AA.
AC Q9FSI3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Beta-amylase.
DE Name=Bmy1;
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;

```



```

RN  SEQUENCE FROM N.A.
RX  MEDLINE=21416776; PubMed=11525070;
RA  Erkkila M.J., Hannu A.;
RT  "Special barley beta-amylase allele in a Finnish landrace line HNS2
RT  with high grain enzyme activity.";
RL  Hereditas 134:91-95(2001).
CC  -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC  polysaccharides so as to remove successive maltose units from the
CC  non-reducing ends of the chains.
CC  -|- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR  EMBL; AJ301645; CAC16789.1; -.
DR  HSSP; P16098; 1B1Y.
DR  GO; GO:0016161; F:beta-amylase activity; IEA.
DR  GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR  GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR  GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR  InterPro; IPR001554; Glyco_hydro_14.
DR  InterPro; IPR001371; Glyco_hydro_14B.
DR  Pfam; PF01373; Glyco_hydro_14; 1.
DR  PRINTS; PR00750; BETAMYLASE.
DR  PRINTS; PR00842; GLHYDLASE14B.
DR  PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR  PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW  Carbohydrate metabolism; Glycosidase; Hydrolase;
KW  Polysaccharide degradation.
SQ  SEQUENCE 535 AA; 59538 MW; 1F30B0865463D6E5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 85
Q9SBH7 PRELIMINARY; PRT; 535 AA.
AC Q9SBH7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Beta-amylase [EC 3.2.1.2].
DE Name-beta-amyl;
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98289114; PubMed=9625721; DOI=10.1104/pp.117.2.679;
RA Erkkila M.J., Leah R., Ahokas H., Cameron-Mills V.;
RT "Allele-dependent barley grain beta-amylase activity.";
RL Plant Physiol. 117:679-695(1998).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -|- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AF061203; AAC67245.1; -.
DR HSSP; P16098; 1B1Y.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001554; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 86
Q7X3X6 PRELIMINARY; PRT; 543 AA.
AC Q7X3X6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative tomatinase Toma.
DE Name-toma;
OS Clavibacter michiganensis subsp. michiganensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=33013;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPB382;
RA Gartenmann K.-H., Graefen I., Zellermann E.-M., Burger A.,
RA Eichenlaub R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393183; AAP57293.1; -.
DR HSSP; P23360; 11IW.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR SMART; SM00633; Glyco_10; 1.
DR SEQUENCE 543 AA; 58175 MW; 8EC284F8A521AED CRC64;

Query Match 0.8%; Score 8; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 LDLITGLV 806
Db 227 LDLITGLV 234

RESULT 87
Q6VIX2 PRELIMINARY; PRT; 546 AA.
AC Q6VIX2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE SFRS protein kinase 2 isoform c.
DE Name=SRPK2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Sha J.H., Zhou Z.M., Li J.M.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR EMBL; AY354201; AAQ63886.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
```

DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0008468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR011009; Kinase like.
 DR InterPro; IPR000719; Prot Kinase.
 DR InterPro; IPR002290; Ser Thr kinase.
 DR InterPro; IPR008271; Ser_thr_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TK; 1.
 DR SMART; SM00219; TyRK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 546 AA; 61049 MW; 5F1DB6805A6CD79C CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 546;
 Best Local Similarity 100.0%; Pred.No. 2.2e+02; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 944 SEKSSSE 951
 Db |||||
 5 SEKSSSE 12
 RESULT 88
 Q6WKW9 PRELIMINARY; PRT; 554 AA.
 AC Q6WKW9;
 DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DE HMg-box protein HMG2L1.
 GN Name=HMG2L1;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8155;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22758517; PubMed=12875653;
 RA Yamada M., Ohkawara B., Ichimura N., Hyodo-Miura J., Urushiyama S.,
 RA Shirakabe K., Shibuya H.;
 RT "Negative regulation of Wnt signalling by HMG2L1, a novel NLK-binding
 protein.";
 RL Genes Cells 8:677-684(2003).
 DR EMBL; AY281348; AAP92714.1; --.
 DR HSP; P07155; IAA.
 DR GO; GO:0000785; C:chromatin; IEA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; F:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000135; Highmobility_12.
 DR InterPro; IPR009071; HMG-box.
 DR Pfam; PF0505; HMG_box; 1.
 DR PRINTS; PR00886; HIGHMOBILITY12.
 DR SMART; SM00398; HMG; 1.
 DR PROSITE; PS01118; HMG_BOX_2; 1.
 SQ SEQUENCE 554 AA; 60909 MW; A2DF64DCA5532F36 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 554;
 Best Local Similarity 100.0%; Pred.No. 2.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 946 KSSSSSE 953
 Db |||||
 430 KSSSSSE 437
 RESULT 89

Q38914 PRELIMINARY; PRT; 555 AA.
 ID Q38914;
 AC Q38914;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE CKC.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Vergani P., Morandini P., Soave C.;
 RL Submitted (DSC-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U44028; AAA86281.1; --.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001471; TP_ERF.
 DR PRINTS; PR00367; ETRSPLEMT.
 DR ProDom; PD001423; TF_ERF; 2.
 DR SMART; SM00380; AP2_2.
 SQ SEQUENCE 555 AA; 61741 MW; A6275D2F082B562D CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 555;
 Best Local Similarity 100.0%; Pred.No. 2.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 LKTVALT 12
 Db |||||
 536 LKTVALT 543
 RESULT 90
 Q42462 PRELIMINARY; PRT; 555 AA.
 ID Q42462;
 AC Q42462;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
 DE ANT (Ovule development protein aintegumenta).
 GN Name=AINTEGUMENTA; Synonyms=AT4937750; At4G37750;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Landsberg erecta; TISSUE=Flower;
 RX MEDLINE=96351414; PubMed=8742707;
 RA Elliott R.C., Betzner A.S., Huttner E., Oakes M.P., Tucker W.Q.,
 RA Gerentes D., Perez P., Smyth D.R.;
 RT "AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic
 roles in ovule development and floral organ growth.";
 RL Plant Cell 8:155-168(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Robben J., Grymonprez B., Volckaert G., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Klucher K.J., Chow H., Reiser L., Fischer R.L.;
 RT "The AINTEGUMENTA Gene of Arabidopsis Required for Ovule and Female
 Gametophyte Development is Related to the Floral Homeotic Gene
 APETALA2.";

Plant Cell 8:0-0(1996).
 RA [5]
 RP SEQUENCE FROM N.A.
 RA Klucher K.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE FROM N.A.
 RA Bewan M., Robben J., Grymonprez B., Volckaert G., Bancroft I.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U41339; AAB17364.1; -;
 DR EMBL; AL161592; CAB80440.1; -;
 DR EMBL; U40256; AAA91040.1; -;
 DR EMBL; AY080706; AAL85024.1; -;
 DR EMBL; AY117207; AAM51282.1; -;
 DR EMBL; AL035709; CAB38923.1; -;
 DR PIR; S71365; S71365.
 DR TRANSFAC; T02639; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001471; TF_ERF.
 DR PRINTS; PR00367; ETRSPLELWNT.
 DR ProDom; PD001423; TF_ERF; 2.
 DR SMART; SM00380; AP2; 2.
 SQ SEQUENCE 555 AA; 61725 MW; BD275D2F082B519B CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 555;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0;
 QY 5 LKTVALTL 12
 Db 536 LKTVALTL 543
 RESULT 91
 Q9P4A3 PRELIMINARY; PRT; 594 AA.
 AC Q9P4A3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Camptothecin resistance conferring protein rcaa.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Emericella.
 OX NCBI_TaxID=162425;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21258791; PubMed=11361337; DOI=10.1007/s004380000411;
 RA Bruschi G.C., de Souza C.C., Fagundes M.R., Dani M.A., Goldman M.H.,

RA Morris N.R., Liu L., Goldman G.H.;
 RT "Sensitivity to camptothecin in Aspergillus nidulans identifies a
 RT novel gene, scaA+, related to the cellular DNA damage response.";
 RL Mol. Genet. Genomics 265:264-275(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Bruschi G.B.C.M., de Souza C.C., Dani M.A.C., Terenzi M.F.,
 RA Goldman M.H.S., Liu L., Morris N.R., Goldman G.H.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF228504; AAF81094.1; -;
 SQ SEQUENCE 594 AA; 67513 MW; 19C4BE8E3DBD4505 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 594;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 891 EGLDIKTI 898
 Db 33 EGLDIKTI 40
 RESULT 92
 Q7VQZ4 PRELIMINARY; PRT; 625 AA.
 AC Q7VQZ4;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Protease IV, a signal peptide peptidase (EC 3.4.21.-).
 GN Name=sppA; OrderedLocusNames=Bfl136;
 OS Candidatus Blochmannia floridanus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; ant endosymbionts; Candidatus Blochmannia.
 OX NCBI_TaxID=203907;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22784745; PubMed=12886019; DOI=10.1073/pnas.1533499100;
 RA Gil R., Silva F.J., Zientz E., Delmotte F., Gonzalez-Candelas F.,
 RA Latorre A., Rausell C., Kamerbeek J., Gadau J., Hoelldobler B.,
 RA van Ham R.C.H.J., Gross R., Moya A.;
 RT "The genome sequence of Blochmannia floridanus: comparative analysis
 RT of reduced genomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:9388-9393(2003).
 DR EMBL; BX248585; CAD83498.1; -;
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0009003; P:signal peptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006465; P:signal peptide processing; IEA.
 DR InterPro; IPR002142; Peptidase_S49.
 DR InterPro; IPR004635; Pept_S49_SppA.
 DR InterPro; IPR004634; Pept_S49_SppA67.
 DR Pfam; PF01343; Peptidase_S49; 2.
 DR PIRSF; PIRSF001217; Protease_4_SppA; 1.
 DR ProDom; PD002897; Peptidase_S49; 1.
 DR TIGRFAMs; TIGR00705; SppA_67K; 1.
 DR TIGRFAMs; TIGR00706; SppA_dcm; 1.
 KW Complete proteome.
 SQ SEQUENCE 625 AA; 71013 MW; D3D4DB1753491AF6 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 625;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 843 KSGKLIYA 850
 Db 141 KSGKLIYA 148
 RESULT 93
 Q9FFV8 PRELIMINARY; PRT; 642 AA.
 AC Q9FFV8;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE Similarity to guanine nucleotide exchange factor (Hypothetical protein
 DE At5g38440; MB518.19) (At5g38440).
 GN Name=At5g38440;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97471969; PubMed=9330910;
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT P1 clones.";
 RL DNA Res. 4:215-230(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Jones T., Carninci P.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
 RA Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
 RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB005231; BAB10154.1; -;
 DR EMBL; AY042523; AAL32601.1; -;
 DR EMBL; BT008789; AAP6227.1; -;
 DR GO; GO:0005851; C:eukaryotic translation initiation factor 2B. . .; IEA.
 DR GO; GO:0005525; F:GTP binding; IEA.
 DR GO; GO:0003743; P:translational initiation factor activity; IEA.
 DR GO; GO:0006413; P:translational initiation; IEA.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR000649; IF-2B.
 DR Pfam; PF01008; IF-2B; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; UNKNOWN 1.
 KW Hypothetical protein.
 SQ SEQUENCE 642 AA; 69495 MW; BBB6507994127893 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 642;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 LTTVSVVT 19
 Db 149 LTTVSVVT 156
 RESULT 94
 Q6ZPU7 PRELIMINARY; PRT; 649 AA.
 AC Q6ZPU7
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DE MKIAA1286 protein (fragment).
 GN Name=MKIAA1286;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

RN SEQUENCE FROM N.A.
 RP TISSUE=Embryonic tail;
 RX PubMed=14621295;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
 RA Saka Y., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT III. the complete nucleotide sequences of 500 mouse KIAA-homologous
 RT cDNAs identified by screening of terminal sequences of cDNA clones
 RT randomly sampled from size-fractionated libraries.";
 RL DNA Res. 10:167-180(2003).
 DR EMBL; AK129322; BAC98132.1; -;
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR000313; PWMP.
 DR Pfam; PF00439; Bromodomain; 1.
 DR Pfam; PF00855; PWMP; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00293; PWMP; 1.
 DR PROSITE; PS50014; BROMODOMAIN 2; 1.
 DR PROSITE; PS50812; PWMP; 1.
 FT NON TER 1
 SQ SEQUENCE 649 AA; 73312 MW; 45F3B0C0EDACF64 CRC64;
 Query Match 0.8%; Score 8; DB 2; Length 649;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 122 LSLNKTVP 129
 Db 295 LSLNKTVP 302
 RESULT 95
 Q6MW74 PRELIMINARY; PRT; 662 AA.
 AC Q6MW74
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DE B1340F09.9 protein.
 GN Name=B1340F09.9;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12447439; DOI=10.1038/nature01183;
 RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
 RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
 RA Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
 RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
 RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
 RA Cai Z., Ren S., Lv G., Zhu G., Tu Y., Jia J., Zhang Y.,
 RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
 RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
 RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
 RA Han B.;
 RT "Sequence and analysis of rice chromosome 4.";
 RL Nature 420:316-320(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
 RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
 RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
 RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
 RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
 RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
 RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
 RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
 RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;

RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
 DR EMBL; BX842608; CAE76071.1; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0005529; F:calcium ion binding; IEA.
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
 DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
 DR GO; GO:0004674; F:transferase activity; IEA.
 DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR011061; Antihomostatic.
 DR InterPro; IPR001881; EGF Ca.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR000719; Kinase_like.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR008271; Ser_thr_pkin_AS.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF07645; EGF CA; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00179; EGF CA; 1.
 DR SMART; SM00220; S_TK; 1.
 DR SMART; SM00219; TyKG; 1.
 DR PROSITE; PS01187; EGF CA; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; EGF-like domain; Kinase; Serine/threonine-protein kinase;
 KW Transferase.
 SQ SEQUENCE 662 AA; 72871 MW; 86E15B4296F87304 CRC64;

 Query Match 0.8%; Score 8; DB 2; Length 662;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 630 LEVASALS 637
 DB 437 LEVASALS 444

 RESULT 96
 Q7XP38 PRELIMINARY; PRT; 673 AA.
 AC Q7XP38;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE OSUNBA0027H09.1 protein.
 GN Name=OSUNBA0027H09.1;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12447439; DOI=10.1038/nature01183;
 RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
 RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
 RA Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu S., Fan D.,
 RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu Z., Guan J.,
 RA Wu M., Zhang R., Zhou B., Chen L., Jin Z., Wang R., Yin H.,
 RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
 RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
 RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
 RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
 RA Han B.;
 RT "Sequence and analysis of rice chromosome 4";
 RL Nature 420:316-320(2002).
 CC -1- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
 DR EMBL; AL662964; CAE03801.2; -.
 DR HSP; P01130; 1H28.
 DR Gramine; Q7XP38; -.

DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
 DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
 DR GO; GO:0004674; F:transferase activity; IEA.
 DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR011061; Antihomostatic.
 DR InterPro; IPR001881; EGF Ca.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR000719; Kinase_like.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkin_AS.
 DR InterPro; IPR008271; Ser_thr_pkinase.
 DR Pfam; PF07645; EGF CA; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00179; EGF CA; 1.
 DR SMART; SM00220; S_TK; 1.
 DR SMART; SM00219; TyKG; 1.
 DR PROSITE; PS01187; EGF CA; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; EGF-like domain; Kinase; Serine/threonine-protein kinase;
 KW Transferase.
 SQ SEQUENCE 673 AA; 74034 MW; 07BEFBA756CAAB9D CRC64;

 Query Match 0.8%; Score 8; DB 2; Length 673;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 630 LEVASALS 637
 DB 437 LEVASALS 444

 RESULT 97
 Q7Z5E6 PRELIMINARY; PRT; 681 AA.
 AC Q7Z5E6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Medulloblastoma antigen MU-MB-20.201 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Medulloblastoma;
 RA Behrends U., Mautner J.M., Muller-Wehrich S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Medulloblastoma;
 RA Behrends U., Mautner J.M., Mautner J.M.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY131226; AAN07912.1; -.
 FT NON TER 1
 SQ SEQUENCE 681 AA; 79727 MW; C6598C3FD2D1425C CRC64;

 Query Match 0.8%; Score 8; DB 2; Length 681;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 793 KRLEKELD 800
 DB 472 KRLEKELD 479

 RESULT 98

O54781 ID 054781 PRELIMINARY; PRT; 681 AA.
 AC O54781
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE SRPK2.
 GN Name=SRPK2;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kuroyanagi N., Onogi H., Wakabayashi T., Hagiwara M.;
 RT "Novel SR-protein-specific kinase, SRPK2, disassembles nuclear
 RT speckles.";
 RL Biochem. Biophys. Res. Commun. 0:0-0(1998).
 CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
 DR EMBL; AB006036; BAA24055.1; -.
 DR F01; JC5929; JC5929.
 DR HSP; Q03656; IHW.
 DR MGD; MGI:1201408; SRPK2.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR001009; Kinase like.
 DR InterPro; IPR000719; Prot Kinase.
 DR Pfam; PF00069; Pkinase; I.
 DR ProDom; PD000001; Prot_kinase; 2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 681 AA; 76740 MW; 1377ADF95C48BC4E CRC64;

 Query Match 0.8%; Score 8; DB 2; Length 681;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 944 SEKSSSSE 951
 Db |||||
 5 SEKSSSSE 12

 RESULT 99
 Q8VCD9 ID Q8VCD9 PRELIMINARY; PRT; 682 AA.
 AC Q8VCD9;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Serine/arginine-rich protein specific kinase 2.
 GN Name=SRPK2;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
 RA Fahney J., Helton E., Kettman M., Maman A.C., Rodriguez S., Sanchez A.,
 RA Whiting M., Maman A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalley D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
 DR EMBL; BC020178; AAH20178.1; -.
 DR F01; JC5929; JC5929.
 DR HSP; Q03656; IHW.
 DR MGD; MGI:1201408; SRPK2.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR Pfam; PF00069; Pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 682 AA; 76884 MW; F519C93A00AC1B9E CRC64;

 Query Match 0.8%; Score 8; DB 2; Length 682;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 944 SEKSSSSE 951
 Db |||||
 5 SEKSSSSE 12

 RESULT 100
 P78362 ID P78362 PRELIMINARY; PRT; 686 AA.
 AC P78362;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Serine kinase SRPK2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98139536; PubMed=9472028; DOI=10.1083/jcb.140.4.737;
 RA Wang H.Y., Lin W., Dyck J.A., Yeakley J.M., Songyang Z., Cantley L.C.,
 RA Fu X.D.;
 RT "SRPK2, a differentially expressed SR protein-specific kinase involved
 RT in mediating the interaction and localization of pre-mRNA splicing
 RT factors in mammalian cells.";
 RL J. Cell Biol. 140:737-750(1998).
 CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
 DR EMBL; U88666; AAC05299.1; -.
 DR HSP; Q03656; IHW.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0004672; F:protein kinase activity; TAS.
 DR GO; GO:0008360; P:RNA splicing; TAS.
 DR GO; GO:0000245; P:spliceosome assembly; TAS.
 DR InterPro; IPR011009; Kinase like.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR008271; Ser_thr_pkin_AS.

DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 686 AA; 77423 MW; 1B9F16E9FEFBAB82 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 686;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSSE 951
|||
Db 5 SEKSSSSE 12

Search completed: August 28, 2005, 11:08:01
Job time : 191 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:49:27 ; Search time 170 Seconds
(without alignments)
2293.262 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 1008

Sequence: 1 MKKHLKTVALLTTSVVVTH.....LGYTSVALLSLTAIKKKY 1008

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 110 summaries

Database : A_Geneseq_16Dec04.*

1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	100.0	1008	5	ABP25813 Streptoco
2	1008	100.0	1008	5	ABP53347 Streptoco
3	1008	100.0	1008	8	ADR83923 S. pyogen
4	76	7.5	1055	4	AAU03612 Group B S
5	76	7.5	1055	5	ABP25812 Streptoco
6	76	7.5	1055	6	ABP56257 Serotype
7	21	2.1	21	5	ABP53348 Streptoco
8	9	0.9	106	4	AAW42172 Human pol
9	9	0.9	235	3	AAQ43601 Arabidops
10	9	0.9	244	3	AAQ43600 Arabidops
11	9	0.9	245	3	AAQ43599 Arabidops
12	9	0.9	279	4	AAQ92626 Human pro
13	9	0.9	279	7	ADJ69129 Human hea
14	9	0.9	340	3	AAQ44240 Arabidops
15	9	0.9	349	3	AAQ44239 Arabidops
16	9	0.9	361	4	AAW40386 Human pol
17	9	0.9	370	3	AAQ44241 Arabidops
18	9	0.9	557	4	AAQ48789 Human pro
19	9	0.9	557	5	ABP97502 Arabidops
20	9	0.9	581	3	AAQ49069 Arabidops
21	9	0.9	583	8	ADJ50969 Human nov
22	9	0.9	590	3	AAQ49068 Arabidops
23	9	0.9	590	5	ABP93671 Herbicida
24	9	0.9	608	7	ABM85378 Mouse pro
25	9	0.9	611	3	AAQ49067 Arabidops

26	9	0.9	692	2	AAW93890	Human HG3
27	9	0.9	883	7	ADB80464	Ovarian c
28	9	0.9	883	7	ADN40012	Cancer/an
29	9	0.9	883	7	ADN39166	Cancer/an
30	9	0.9	907	2	AAW93889	Human HG3
31	9	0.9	907	3	AAW90682	Human G p
32	9	0.9	907	3	AAW90687	Human mut
33	9	0.9	907	6	ABP81968	Human G p
34	9	0.9	907	6	ABO06467	Human G p
35	9	0.9	907	7	ADC22797	Human G p
36	9	0.9	907	7	ADC22783	Human G p
37	9	0.9	907	7	ADE59150	Human Pro
38	9	0.9	907	7	ADE59153	Human Pro
39	9	0.9	907	7	ADG42629	Human G p
40	9	0.9	907	7	ADG42628	Human G p
41	9	0.9	907	7	ADH14270	Mutated h
42	9	0.9	907	7	ADH14256	Human HG3
43	9	0.9	907	7	ADN40013	Cancer/an
44	9	0.9	907	7	ADN39531	Cancer/an
45	9	0.9	907	7	ADN39628	Cancer/an
46	9	0.9	907	8	ADO29408	Human GPC
47	9	0.9	907	8	ADQ80369	G protein
48	9	0.9	907	8	ADR67868	Human HG3
49	9	0.9	1145	7	ADF70480	Orphan re
50	8	0.8	147	4	ABG03518	Novel hum
51	8	0.8	196	7	ADG88441	Arabidops
52	8	0.8	240	7	ADM26567	Hyperther
53	8	0.8	264	7	ADG88453	Arabidops
54	8	0.8	268	5	ABH05713	Human cel
55	8	0.8	275	7	ADG88454	Arabidops
56	8	0.8	277	7	ADG88440	Arabidops
57	8	0.8	284	3	AAI18240	Plasmodi
58	8	0.8	308	7	ADG88439	Arabidops
59	8	0.8	319	2	AAV03187	Rat Acid
60	8	0.8	350	2	AAV22200	S. frugip
61	8	0.8	356	6	ABU17197	Protein e
62	8	0.8	364	6	ADA33080	Acinetoba
63	8	0.8	411	5	ABP73447	Candida a
64	8	0.8	458	8	ADS28368	Bacterial
65	8	0.8	473	7	ADG88438	Arabidops
66	8	0.8	480	7	ABC08227	Rice prot
67	8	0.8	483	4	AAW92603	Human pro
68	8	0.8	488	8	ADJ48708	Oil-assoc
69	8	0.8	488	8	ADJ49363	Oil-assoc
70	8	0.8	488	8	ADJ48727	Oil-assoc
71	8	0.8	488	8	ADJ49362	Oil-assoc
72	8	0.8	496	7	ADC07838	Rice prot
73	8	0.8	500	8	ADJ50282	Oil-assoc
74	8	0.8	503	8	ADJ49407	Oil-assoc
75	8	0.8	503	8	ADJ48856	Oil-assoc
76	8	0.8	503	8	ADJ50184	Oil-assoc
77	8	0.8	519	4	AAW94834	Human pro
78	8	0.8	523	2	AAW67006	Recombina
79	8	0.8	526	2	AAW88327	Beta-amyl
80	8	0.8	527	8	ADJ48855	Oil-assoc
81	8	0.8	528	3	AAQ31382	Arabidops
82	8	0.8	529	3	AAQ31381	Arabidops
83	8	0.8	530	2	AAW67005	Barley be
84	8	0.8	531	2	AAW97613	Beta-amyl
85	8	0.8	533	8	ADJ49017	Oil-assoc
86	8	0.8	533	8	ADJ49016	Oil-assoc
87	8	0.8	535	2	AAW04261	Beta-amyl
88	8	0.8	535	2	AAW15746	Mutated b
89	8	0.8	535	8	ADJ49120	Oil-assoc
90	8	0.8	535	8	ADJ50393	Oil-assoc
91	8	0.8	535	8	ADJ50119	Oil-assoc
92	8	0.8	535	8	ADJ49037	Oil-assoc
93	8	0.8	535	8	ADJ50182	Oil-assoc
94	8	0.8	555	3	AAW07724	An Arabid
95	8	0.8	555	3	AAW07724	An Arabid
96	8	0.8	555	6	ABR40832	Arabidops
97	8	0.8	555	6	ABR40830	Arabidops
98	8	0.8	555	7	ADG25136	Thalecres

99 8 0.8 555 7 ADG88451 Arabidops
 100 8 0.8 555 7 ADG88431 Arabidops
 101 8 0.8 555 7 ADG88450 Arabidops
 102 8 0.8 555 8 ADO61539 Transcrip
 103 8 0.8 649 5 ABP69424
 104 8 0.8 660 4 ABG30817
 105 8 0.8 660 5 ABG38772
 106 8 0.8 681 7 ADE59481
 107 8 0.8 686 5 AAU80373 Human cel
 108 8 0.8 686 6 AAE34823 Protein #
 109 ADO19347 Human PRO
 110 8 0.8 688 2 AAY27053 Human pro

ALIGNMENTS

RESULT 1
 ABP25813
 ID ABP25813 standard; protein; 1008 AA.
 XX
 AC ABP25813;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 802.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus pyogenes.
 XX
 WN WO200234771-A2.
 PN
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB004789.
 XX
 PR 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;
 PI Tettelin H;
 XX
 DR WPI: 2002-352536/38.
 DR N-PSDB; ABN66444.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX
 PS Claim 1; Page 3231; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX
 SQ Sequence 1008 AA;
 Query Match 100.0%; Score 1008; DB 5; Length 1008;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MKGHLKVALTLTVSVVTHNQEVFSLVKGPILKQTQASSISISGADYAESSGSKLKLINE 60
 Db 1 MKGHLKVALTLTVSVVTHNQEVFSLVKGPILKQTQASSISISGADYAESSGSKLKLINE 60
 Qy 61 TSGPVDVDTVDFSDKETTPEKIKDNLAKGPREQLKAVTENTSEKQITSGSOLESKKE 120
 Db 61 TSGPVDVDTVDFSDKETTPEKIKDNLAKGPREQLKAVTENTSEKQITSGSOLESKKE 120
 Qy 121 SLSLNTKVPSTSNWEICDFITKGNLTVLGLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
 Db 121 SLSLNTKVPSTSNWEICDFITKGNLTVLGLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
 Qy 181 FAFTPDKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFNLSLLKKVTIPTGYKHIGQ 240
 Db 181 FAFTPDKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFNLSLLKKVTIPTGYKHIGQ 240
 Qy 241 DAFVDNKNIAEVLNPESLETISDYAFALALKOITDLPDNLKAIGELAFPDNQITGKLSLP 300
 Db 241 DAFVDNKNIAEVLNPESLETISDYAFALALKOITDLPDNLKAIGELAFPDNQITGKLSLP 300
 Qy 301 RQLMLRAERAFKSNHIKTIEFRGNSLKVIGEASQDNDLSQMLPDGLIEKIESAFTGNP 360
 Db 301 RQLMLRAERAFKSNHIKTIEFRGNSLKVIGEASQDNDLSQMLPDGLIEKIESAFTGNP 360
 Qy 361 GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
 Db 361 GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
 Qy 421 GFSNKGLOKVKRKNLEIPKOHNGVTITEIGDGNFRVNDPONTLRKYVDLEEVKLPSITR 480
 Db 421 GFSNKGLOKVKRKNLEIPKOHNGVTITEIGDGNFRVNDPONTLRKYVDLEEVKLPSITR 480
 Qy 481 KIGAFAFQSNLKSFEASDDLEEKEGAFMNNRIETLELKKDLVTIGDAAFHINHIYAIV 540
 Db 481 KIGAFAFQSNLKSFEASDDLEEKEGAFMNNRIETLELKKDLVTIGDAAFHINHIYAIV 540
 Qy 541 LPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSEKQLTETIPVQAF 600
 Db 541 LPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSEKQLTETIPVQAF 600
 Qy 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSHIAFNALDDNDGDGFONKVVVK 660
 Db 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSHIAFNALDDNDGDGFONKVVVK 660
 Qy 661 THNSYALADGEHFIVDPDKLSSTIVDLKILKIEGLDYSTLTQTQTQFRDMMTAGKA 720
 Db 661 THNSYALADGEHFIVDPDKLSSTIVDLKILKIEGLDYSTLTQTQTQFRDMMTAGKA 720
 Qy 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAKEALVTKKATKNGQLLERSINKAVL 780
 Db 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAKEALVTKKATKNGQLLERSINKAVL 780
 Qy 781 AYNNSAIKKNVKELEKELDLITGLVSGKGLAQAATWVGYYLAKTLPPLPEYYIGLVNY 840
 Db 781 AYNNSAIKKNVKELEKELDLITGLVSGKGLAQAATWVGYYLAKTLPPLPEYYIGLVNY 840
 Qy 841 FDKSGKLIYALDMSDTIGEGOKDAYGNPILNVDSNEGYYHALAVATLADYEGLDIKTILN 900
 Db 841 FDKSGKLIYALDMSDTIGEGOKDAYGNPILNVDSNEGYYHALAVATLADYEGLDIKTILN 900
 Qy 901 SKLSQLSIROPVTAAYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSSSSANSKDRGL 960
 Db 901 SKLSQLSIROPVTAAYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSSSSANSKDRGL 960

Qy	961	QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLTAIKKKY	1008
Db	961	QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLTAIKKKY	1008

RESULT 2

ABP53347
ID ABP53347 standard; protein; 1008 AA.

AC ABP53347;

DT 18-NOV-2002 (first entry)

DE Streptococcus pyogenes BVH-P7 protein SEQ ID NO:2.

Streptococcus pyogenes; BVH-P7; antigen; group A Streptococcus; GAS; bacterial pathogen; vaccine; antibacterial; gene therapy; pharyngitis; streptococcal infection; erysipelas; impetigo; scarlet fever; infection; invasive disease; bacteraemia; necrotising fasciitis; toxic shock.

OS *Streptococcus pyogenes*.

AA	FH	Key	Location/Qualifiers
----	----	-----	---------------------

FT	Peptide	1. .21
FT	Peptide	1. .21

FT /label= signal

FT Protein 22. .1008

FT /label= mature_BVH-p7

PN WO20026650-A2.

PD 29-AUG-2002.

21-FEB-2002; 2002WO-CA000207.

PR 21-FEB-2001; 2001US-0269840P.

PA (SHIR-) SHIRE BIOCHEM INC.

PI Martin D, Rioux S, Brodeur BR, Hamel J, Rheault P;

DR WPI; 2002-674948/72.

DR N-PSDB; ABQ81821.

PT New polypeptide useful as a vaccine component for preventing, treating or
PT diagnosing *Streptococcus pyogenes* infections, e.g. pharyngitis,
PT erysipelas, impetigo, scarlet fever, bacteremia, necrotizing fasciitis or
PT toxic shock.

PS Claim 17; Fig 2; 52pp; English.

The present sequence represents *Streptococcus pyogenes* BVH-P7 protein (I). (I) has antibacterial activity and can be used in vaccines and gene therapy. The *Streptococcus pyogenes* BVH-P7 polypeptide is useful as a vaccine component for preventing, treating and/or diagnosing streptococcal infections, such as pharyngitis, erysipelas, impetigo, scarlet fever, and invasive diseases such as bacteraemia and necrotising fasciitis, or toxic shock. A composition comprising the BVH-P7 polypeptide is useful in the manufacture of a medicament for the prophylactic or therapeutic treatment of streptococcal infection. The BVH-P7 polynucleotide may be used in designing probes for the detection of *Streptococcus* in biological samples. The BVH-P7 polypeptide may also be used as immunogens for the production of antibodies against streptococcal infections.

Sequence 1008 AA;

Query Match 100.0%; Score 1008; DB 5; Length 1008;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKKHLKTVALTLTTVSVVTNHQEVFSLVKEPILKQTQASSISGADYAESSGSKLKINE 60

Db 1 MKKHLKTVALTLTTVSVVTHNQEVFSLVKEPILKQTOASSISGADYAESSGKSKLKINE 60

RESULT 3

ADR83923

ID ADR83923 standard; protein; 1008 AA.

AC ADR83923;

DT 02-DEC-2004 (first entry)

Qy	61	TSGPVDVTDLFSDKRTTPEKIKDNLAGPREQELKAVTENTESEKQITSGSQLEQSK	120
Db	61	TSGPVDVTDLFSDKRTTPEKIKDNLAGPREQELKAVTENTESEKQITSGSQLEQSK	120
Qy	121	SLSLNKTVPSTSNWEICDFTIKGNTLVGLSGSGVEKLSQTDHLVLPQAAADGTQLIQVAS	180
Db	121	SLSLNKTVPSTSNWEICDFTIKGNTLVGLSGSGVEKLSQTDHLVLPQAAADGTQLIQVAS	180
Qy	181	FAFTPDKKTAAIEVYTSRAGENGEBISOLDVDGKEIINEGEVFNYSYLLKKVYIPTGYKHIGQ	240
Db	181	FAFTPDKKTAAIEVYTSRAGENGEBISOLDVDGKEIINEGEVFNYSYLLKKVYIPTGYKHIGQ	240
Qy	241	DAFVDNKNIAEVNLPESILETSDYAFPAHALKQIDLPNLKAIAGELAFDNOITGKLSIP	300
Db	241	DAFVDNKNIAEVNLPESILETSDYAFPAHALKQIDLPNLKAIAGELAFDNOITGKLSIP	300
Qy	301	QQLMRLAERAFKSNHIIKTIFPRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEFTGNP	360
Db	301	QQLMRLAERAFKSNHIIKTIFPRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEFTGNP	360
Qy	361	GDHYNRVVLWTGSGKNPGLATENTVYVNPDKSLWQESPEIDYTKWLEEDPTYQKNSVT	420
Db	361	GDHYNRVVLWTGSGKNPGLATENTVYVNPDKSLWQESPEIDYTKWLEEDPTYQKNSVT	420
Qy	421	GFNSKGLQVKRNKNLBEIPKOHNGVTITEIGDNAPRVNDFQNKTKRYDLEBVKLPSTIR	480
Db	421	GFNSKGLQVKRNKNLBEIPKOHNGVTITEIGDNAPRVNDFQNKTKRYDLEBVKLPSTIR	480
Qy	481	KIGAFAFQSNLKSFEASDDLEETKEGAFMNNRLETLELKOKLVTIGDAAAPHNIHAIYV	540
Db	481	KIGAFAFQSNLKSFEASDDLEETKEGAFMNNRLETLELKOKLVTIGDAAAPHNIHAIYV	540
Qy	541	LPESVQBIGRSAFRQNGANNLIFMGSKVYKTIGEMAFLSNRLEHLDSLQKQLTEIPVQAF	600
Db	541	LPESVQBIGRSAFRQNGANNLIFMGSKVYKTIGEMAFLSNRLEHLDSLQKQLTEIPVQAF	600
Qy	601	SDNALKEVLLPASLKTITREAFKQNLKQLEVASALSHIAFNALDDNDGDEQFNKVVVK	660
Db	601	SDNALKEVLLPASLKTITREAFKQNLKQLEVASALSHIAFNALDDNDGDEQFNKVVVK	660
Qy	661	THNSYALADGEHPIVDPDKLSSTIVLEKILKILIEGLDYSTLRQTTOTQPRDMTTAGKA	720
Db	661	THNSYALADGEHPIVDPDKLSSTIVLEKILKILIEGLDYSTLRQTTOTQPRDMTTAGKA	720
Qy	721	LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAAEKALVTKKATKNGQLLERSINKAVL	780
Db	721	LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAAEKALVTKKATKNGQLLERSINKAVL	780
Qy	781	AYNNSAIKKNVKELEKELDLLTGLVFGKGPLAQATWVGQVYLLKTPILPEYIYGLNYY	840
Db	781	AYNNSAIKKNVKELEKELDLLTGLVFGKGPLAQATWVGQVYLLKTPILPEYIYGLNYY	840
Qy	841	FDKSGKLIYALDMSDDTIGEGQKQAYGNPILNVDSDEGNYHALAVATLADYEGLDIKTILN	900
Db	841	FDKSGKLIYALDMSDDTIGEGQKQAYGNPILNVDSDEGNYHALAVATLADYEGLDIKTILN	900
Qy	901	SKLSQLTISRQVPTAAVYHRAGIFQAIQNAAAEAEQLLPKPGTHSEKSSSSSESANSKDRGL	960
Db	901	SKLSQLTISRQVPTAAVYHRAGIFQAIQNAAAEAEQLLPKPGTHSEKSSSSSESANSKDRGL	960
Qy	961	QSNPKNTRGRHSAILPRTGSGKSPVYIGLVYTSVALLSLITAIKKKKY 1008	
Db	961	QSNPKNTRGRHSAILPRTGSGKSPVYIGLVYTSVALLSLITAIKKKKY 1008	

RESULT 3

ADR83923

ID ADR83923 standard; protein; 1008 AA.

AC ADR83923;

DT 02-DEC-2004 (first entry)

XX S. pyogenes hyperimmune system reactive antigen Spy0843.
 XX hyperimmune serum reactive antigen; vaccine; anticaline.
 XX Streptococcus pyogenes.
 XX WO2004078907-A2.
 XX 16-SEP-2004.
 XX 02-MAR-2004; 2004WO-EP002087.
 XX 04-MAR-2003; 2003EP-00450061.
 XX (INTE-) INTERCELL AG.
 XX Meinke A, Nagy E, Winkler B, Gelbmann D;
 XX WPI: 2004-653698/63.
 XX N-PSDB; ADR83773.
 XX New isolated nucleic acid molecules encoding hyperimmune serum-reactive
 PT antigens from Streptococcus pyogenes, useful for diagnosing, preventing
 PT and treating S. pyogenes infections.
 XX Claim 14; SEQ ID NO 191; 145pp; English.
 XX This invention describes a novel nucleic acid molecule encoding a
 CC hyperimmune serum reactive antigen or its fragment from Streptococcus
 CC pyogenes. The nucleic acid molecule or hyperimmune serum-reactive antigen
 CC or its fragment are useful for the manufacture of a pharmaceutical
 CC preparation, especially a vaccine, against S. pyogenes infection. In
 CC addition, the hyperimmune serum reactive antigen or fragment is used for
 CC the isolation and/or purification and/or identification of an interaction
 CC partner of the hyperimmune serum reactive antigen or its fragment, for
 CC the generation of a peptide (e.g. anticalines) binding to the antigen or
 CC fragment, or for the manufacture of a functional nucleic acid selected
 CC from aptamers and spiegelmers. The nucleic acid molecule may also be used
 CC for the manufacture of functional ribonucleic acids, such as ribozymes,
 CC antisense nucleic acids and siRNA. ADR83773-ADR84189 represent S.
 CC pyogenes hyperimmune serum reactive antigens, fragments and the encoding
 CC polynucleotide described in the invention.
 XX Sequence 1008 AA;

Query Match 100.0%; Score 1008; DB 8; Length 1008;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKELKTVALTUTTVSVTHNQEVPSLVKPEILKQTOASSISGADYAESGSKLKINE 60
 DB 1 MKKELKTVALTUTTVSVTHNQEVPSLVKPEILKQTOASSISGADYAESGSKLKINE 60

QY 61 TSGPVDVTDLFSPDKRTTPEKIKONLAKGPREQELKAVTENTESEKQITSGSQLEQSK 120
 DB 61 TSGPVDVTDLFSPDKRTTPEKIKONLAKGPREQELKAVTENTESEKQITSGSQLEQSK 120

QY 121 SLSLNTKVPSTSNWICDPIFKGNTLVGLSKGVEKLSQTDHLVLPSSQAADGTQLIQVAS 180
 DB 121 SLSLNTKVPSTSNWICDPIFKGNTLVGLSKGVEKLSQTDHLVLPSSQAADGTQLIQVAS 180

QY 181 FAFTPDKKTATIAEYTSRAGENGEISQLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
 DB 181 FAFTPDKKTATIAEYTSRAGENGEISQLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240

QY 241 DAFVONKNAEYNLPESLETISDYAFALHAKQIDLPNLKAIKAGELAFDQITKLSLP 300
 DB 241 DAFVONKNAEYNLPESLETISDYAFALHAKQIDLPNLKAIKAGELAFDQITKLSLP 300

QY 301 RQMLAERAPKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360
 DB 301 RQMLAERAPKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360

QY 361 GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
 DB 361 GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

QY 421 GFSNKGKQKVRKNKLEIPKQHNQVTTIEIGDNAFRNVDFONKTLRKYDLEEVKLPSTIR 480
 DB 421 GFSNKGKQKVRKNKLEIPKQHNQVTTIEIGDNAFRNVDFONKTLRKYDLEEVKLPSTIR 480

QY 481 KIGAPAFQSNNLKSPFASDDLEEIKEGAFMNNRIETLELKDCLVTIGDAAAPHINHIAIV 540
 DB 481 KIGAPAFQSNNLKSPFASDDLEEIKEGAFMNNRIETLELKDCLVTIGDAAAPHINHIAIV 540

QY 541 LPESVQIEGRSAFRONGANNLI FMGSKVKTIGENAFPLSNRLEHLDLSEKQKLTPIVQAF 600
 DB 541 LPESVQIEGRSAFRONGANNLI FMGSKVKTIGENAFPLSNRLEHLDLSEKQKLTPIVQAF 600

QY 601 SDNALKEVLLPASLKTIREERAFKKNHLKQLEVASALSHIAFNALDDNDGDQFONKVVVK 660
 DB 601 SDNALKEVLLPASLKTIREERAFKKNHLKQLEVASALSHIAFNALDDNDGDQFONKVVVK 660

QY 661 THNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLTROTQTQFPRDMTTAGKA 720
 DB 661 THNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLTROTQTQFPRDMTTAGKA 720

QY 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780
 DB 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780

QY 781 AYNNSAIKKNVKKLEKELDLTGLVSGKGLAQATWVGYYLLKTPLPPEYIIGLVNY 840
 DB 781 AYNNSAIKKNVKKLEKELDLTGLVSGKGLAQATWVGYYLLKTPLPPEYIIGLVNY 840

QY 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDSNDEGYHALAVATLADYBGLDIKTILN 900
 DB 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDSNDEGYHALAVATLADYBGLDIKTILN 900

QY 901 SKLSQLSIRQVPTAAVHRAGIFQAIQNAABAEQLLPKPTHSEKSSSSSANSKDRGL 960
 DB 901 SKLSQLSIRQVPTAAVHRAGIFQAIQNAABAEQLLPKPTHSEKSSSSSANSKDRGL 960

QY 961 QSNPKTNRGRHSAILPRTGSKGSPVYGLVTSVALLSLITAIKKKKY 1008
 DB 961 QSNPKTNRGRHSAILPRTGSKGSPVYGLVTSVALLSLITAIKKKKY 1008

RESULT 4
 AAU03612
 ID AAU03612 standard; protein; 1055 AA.
 XX
 AC AAU03612;
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Group B Streptococcus antigenic protein, ID-87.
 XX
 KW Group B Streptococcus; encapsulated bacterium; therapeutic; sepsis;
 KW meningitis; neonate; antigenic; vaccine; infection; genital tract;
 KW capsid polysaccharide vaccination.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200132882-A2.
 XX
 PD 10-MAY-2001.
 XX
 PF 07-SEP-2000; 2000WO-GB003437.
 XX
 PR 07-SEP-1999; 99GB-00021125.
 XX
 PA (MICR-) MICROBIAL TECHNIQS LTD.
 XX
 PI Le Page RWP, Wells JM, Hanniffy SB;

XX WPI; 2001-316444/33.
 DR N-PSDB; AAS07029.
 XX
 PT New polypeptides derived from Streptococcus agalactiae are useful to
 PT provide detection of, and vaccination against, Group B Streptococcus
 PT infections, particularly to prevent infection in neonates.
 XX
 PS Claim 1; Fig 1; 178pp; English.
 XX
 CC AAU03601-AAU03722 represent Group B Streptococcus (Streptococcus
 CC agalactiae) amino acid sequences of the invention. S. agalactiae is an
 CC encapsulated bacterium which is a major pathogen of humans causing sepsis
 CC and meningitis in neonates as well as adults. The S. agalactiae antigenic
 CC polypeptides are used to vaccinate against Group B Streptococcus
 CC infections, particularly to prevent infection in new born children
 CC arising from the maternal genital tract. An immunogenic composition is
 CC useful in the preparation of a medicament for the treatment or
 CC prophylaxis of Group B Streptococcus infection. The invention does not
 CC have the disadvantages of varied response rate associated with prior art
 CC capsid polysaccharide vaccination against Group B Streptococcus
 XX
 SQ Sequence 1055 AA;
 Query Match 7.5%; Score 76; DB 4; Length 1055;
 Best Local Similarity 100.0%; Pred. No. 6.6e-65;
 Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 805 LVEGKGLAQTAVQGVYLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDTIGGQKDA 864
 DB 806 LVEGKGLAQTAVQGVYLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDTIGGQKDA 865
 QY 865 YGNPILNVDEDNDEGYH 880
 DB 866 YGNPILNVDEDNDEGYH 881
 RESULT 5
 ABP25812
 ID ABP25812 standard; protein; 1055 AA.
 XX
 AC ABP25812;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 800.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB004789.
 XX
 PR 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Massignani V, Margarit Y Rosl, Grandi G, Fraser C;
 PI Tettelin H;
 XX
 DR WPI; 2002-352536/38.
 DR N-PSDB; ABN66443.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection or

PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX
 PS Claim 1; Page 3230-3231; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX
 SQ Sequence 1055 AA;
 Query Match 7.5%; Score 76; DB 5; Length 1055;
 Best Local Similarity 100.0%; Pred. No. 6.6e-65;
 Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 805 LVEGKGLAQTAVQGVYLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDTIGGQKDA 864
 DB 806 LVEGKGLAQTAVQGVYLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDTIGGQKDA 865
 QY 865 YGNPILNVDEDNDEGYH 880
 DB 866 YGNPILNVDEDNDEGYH 881
 RESULT 6
 ABP56257
 ID ABP56257 standard; protein; 1055 AA.
 XX
 AC ABP56257;
 XX
 DT 28-MAR-2003 (first entry)
 XX
 DE Serotype III group B Streptococcus strain COH1 BVH-A4 SEQ ID NO:2.
 XX
 KW Serotype III group B streptococcus strain COH1; BVH-A4; streptococcus;
 KW antibiotic; immunostimulant; vaccine; bacterial infection; sepsis;
 KW meningitis; pneumonia; cellulitis; osteomyelitis; septic arthritis;
 KW endocarditis; epiglottitis; osteomyelitis; amniotitis; endometritis;
 KW cellulitis; fasciitis; bacteraemia; urosepsis; peritonitis; emphysema;
 KW mastitis; streptococcal infection.
 XX
 OS Streptococcus sp.
 XX
 FH Key Location/Qualifiers
 FH Peptide 1..22
 FT /label= signal
 FT Protein 23..1055
 FT /label= BVH-A4
 XX
 PN WO200288178-A2.
 XX
 PD 07-NOV-2002.
 XX
 PF 02-MAY-2002; 2002WO-CA000664.
 XX
 PR 02-MAY-2001; 2001US-0287712P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.

PI Martin D, Hamel J, Brodeur BR, Rioux S, Boyer M;
 XX
 DR WPI; 2003-120461/11.
 DR N-PSDB; ABZ21973.
 XX
 PT New BVH-A4 proteins and genes from serotype III Group B streptococcus,
 PT useful for treating or preventing streptococcal infection in infants,
 PT pregnant women, non-pregnant adults (e.g. pneumonia), or members of dairy
 PT herd (mastitis).
 XX
 PS Claim 18; Fig 2; 60pp; English.
 XX
 CC The present sequence represents a BVH-A4 protein from serotype III Group
 CC B streptococcus (GBS) strain COH1 (designated GBS-BVH-A4) (I). (I) has
 CC antibiotic and immunostimulant activities, and can be used in polypeptide
 CC therapy and in vaccine production. (I) can be used for the therapeutic or
 CC prophylactic treatment of GBS bacterial infection in a host susceptible
 CC to GBS infection. In particular, (I) is useful for treating or preventing
 CC GBS infection in a neonate or infant (e.g. sepsis, meningitis, pneumonia,
 CC cellulitis, osteomyelitis, septic arthritis, endocarditis or
 CC epiglottitis), in a pregnant woman (e.g. mild urinary tract infection to
 CC life-threatening sepsis and meningitis, osteomyelitis, endocarditis,
 CC amniotitis, endometritis, wound infection (post-caesarean or post-
 CC episiotomy), cellulitis or fasciitis), in a non-pregnant adult (e.g.
 CC bacteraemia, skin or soft tissue infection, pneumonia, urosepsis,
 CC endocarditis, peritonitis, meningitis or empysema), or in a member of
 CC dairy herd (e.g. mastitis). A composition comprising (I) or (I) can also
 CC be useful for treating or preventing streptococcal infection
 XX
 SQ Sequence 1055 AA;
 Query Match 7.5%; Score 76; DB 6; Length 1055;
 Best Local Similarity 100.0%; Pred. No. 6.6e-65;
 Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 805 LVEGKGLAQATMVGVLLKTPLEPYIGLVNVPYFDKSGKLIYALDMSDTIGSGQKDA 864
 DB 806 LVEGKGLAQATMVGVLLKTPLEPYIGLVNVPYFDKSGKLIYALDMSDTIGSGQKDA 865
 QY 865 YGNPILNVDNEGYH 880
 DB 866 YGNPILNVDNEGYH 881
 RESULT 7
 ID ABP53348
 XX ABP53348 standard; peptide; 21 AA.
 AC ABP53348;
 XX 18-NOV-2002 (first entry)
 DE Streptococcus pyogenes BVH-P7 signal peptide.
 KW Streptococcus pyogenes; BVH-P7; antigen; group A Streptococcus; GAS;
 KW bacterial pathogen; vaccine; antibacterial; gene therapy; pharyngitis;
 KW streptococcal infection; erysipelas; impetigo; scarlet fever; infection;
 KW invasive disease; bacteraemia; necrotizing fasciitis; toxic shock.
 XX
 OS Streptococcus pyogenes.
 XX WO200266650-A2.
 PN 29-AUG-2002.
 PD 21-FEB-2002; 2002WO-CA000207.
 PF 21-FEB-2001; 2001US-0269840P.
 PR (SHIR-) SHIRE BIOCHEM INC.
 PA Martin D, Rioux S, Brodeur BR, Hamel J, Rheault P;
 PI
 XX

DR WPI; 2002-674948/72.
 XX
 PT New polypeptide useful as a vaccine component for preventing, treating or
 PT diagnosing Streptococcus pyogenes infections, e.g. pharyngitis,
 PT erysipelas, impetigo, scarlet fever, bacteraemia, necrotizing fasciitis or
 PT toxic shock.
 XX
 PS Example 1; Page 25; 52pp; English.
 XX
 CC The present invention describes Streptococcus pyogenes BVH-P7 protein
 CC (I). (I) has antibacterial activity and can be used in vaccines and gene
 CC therapy. The Streptococcus pyogenes BVH-P7 polypeptide is useful as a
 CC vaccine component for preventing, treating and/or diagnosing
 CC streptococcal infections, such as pharyngitis, erysipelas, impetigo,
 CC scarlet fever, and invasive diseases such as bacteraemia and necrotizing
 CC fasciitis, or toxic shock. A composition comprising the BVH-P7
 CC polypeptide is useful in the manufacture of a medicament for the
 CC prophylactic or therapeutic treatment of streptococcal infection. The BVH
 CC -P7 polynucleotide may be used in designing probes for the detection of
 CC Streptococcus in biological samples. The BVH-P7 polypeptide may also be
 CC used as immunogens for the production of antibodies against streptococcal
 CC infections. The present sequence represents the signal peptide of S.
 CC pyogenes BVH-P7, which is given in an example from the present invention
 XX
 SQ Sequence 21 AA;
 Query Match 2.1%; Score 21; DB 5; Length 21;
 Best Local Similarity 100.0%; Pred. No. 2.2e-12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKGHLKTVALTLTTVSVVTHN 21
 DB 1 MKGHLKTVALTLTTVSVVTHN 21
 RESULT 8
 ID AAM42172
 XX AAM42172 standard; protein; 106 AA.
 AC AAM42172;
 XX 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 7103.
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX WO200153312-A1.
 PN 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US034263.
 PF 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00652191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB; AA161328.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Example 2; SEQ ID NO 7103; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AA38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
XX Sequence 106 AA;
SQ

Query Match 0.9%; Score 9; DB 4; Length 106;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952
Db |||||
36 SEKSSSES 44

RESULT 9
AAG43601
ID AAG43601 standard; protein; 235 AA.
XX
AC AAG43601;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 54515.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.

23-APR-1999; 99US-0130891P.
28-APR-1999; 99US-0131449P.
30-APR-1999; 99US-0132048P.
30-APR-1999; 99US-0132407P.
04-MAY-1999; 99US-0132484P.
05-MAY-1999; 99US-0132485P.
06-MAY-1999; 99US-0132486P.
06-MAY-1999; 99US-0132487P.
07-MAY-1999; 99US-0132863P.
11-MAY-1999; 99US-0134256P.
14-MAY-1999; 99US-0134218P.
14-MAY-1999; 99US-0134219P.
14-MAY-1999; 99US-0134221P.
18-MAY-1999; 99US-0134370P.
18-MAY-1999; 99US-0134769P.
19-MAY-1999; 99US-0134941P.
20-MAY-1999; 99US-0135124P.
21-MAY-1999; 99US-0135353P.
24-MAY-1999; 99US-0135629P.
25-MAY-1999; 99US-0136021P.
27-MAY-1999; 99US-0136392P.
28-MAY-1999; 99US-0136782P.
01-JUN-1999; 99US-0137222P.
03-JUN-1999; 99US-0137528P.
04-JUN-1999; 99US-0137502P.
07-JUN-1999; 99US-0137724P.
08-JUN-1999; 99US-0138094P.
10-JUN-1999; 99US-0138540P.
10-JUN-1999; 99US-0138847P.
14-JUN-1999; 99US-0139119P.
16-JUN-1999; 99US-0139452P.
16-JUN-1999; 99US-0139453P.
17-JUN-1999; 99US-0139492P.
18-JUN-1999; 99US-0139454P.
18-JUN-1999; 99US-0139455P.
18-JUN-1999; 99US-0139456P.
18-JUN-1999; 99US-0139457P.
18-JUN-1999; 99US-0139458P.
18-JUN-1999; 99US-0139459P.
18-JUN-1999; 99US-0139460P.
18-JUN-1999; 99US-0139461P.
18-JUN-1999; 99US-0139462P.
18-JUN-1999; 99US-0139463P.
18-JUN-1999; 99US-0139750P.
18-JUN-1999; 99US-0139763P.
21-JUN-1999; 99US-0139817P.
22-JUN-1999; 99US-0139899P.
23-JUN-1999; 99US-0140353P.
23-JUN-1999; 99US-0140354P.
24-JUN-1999; 99US-0140695P.
28-JUN-1999; 99US-0140823P.
29-JUN-1999; 99US-0140991P.
30-JUN-1999; 99US-0141287P.
01-JUL-1999; 99US-0141842P.
01-JUL-1999; 99US-0142154P.
02-JUL-1999; 99US-0142055P.
06-JUL-1999; 99US-0142390P.
08-JUL-1999; 99US-0142803P.
09-JUL-1999; 99US-0142920P.
12-JUL-1999; 99US-0142977P.
13-JUL-1999; 99US-0143542P.
14-JUL-1999; 99US-0143624P.
15-JUL-1999; 99US-0144005P.
16-JUL-1999; 99US-0144085P.
16-JUL-1999; 99US-0144086P.
19-JUL-1999; 99US-0144325P.
19-JUL-1999; 99US-0144331P.
19-JUL-1999; 99US-0144332P.
19-JUL-1999; 99US-0144333P.
19-JUL-1999; 99US-0144334P.
19-JUL-1999; 99US-0144335P.
20-JUL-1999; 99US-0144352P.
20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0158294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.

PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160788P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 235;
Best Local Similarity 100.0%; Pred.No.13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAARQQL 937

Db 123 AAARQQL 131

RESULT 10

AAAG3600

ID AAAG3600 standard; protein; 244 AA.

AC AAG43600;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 54514.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0130891P.

PR 30-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.


```
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 244;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAEAEQLL 937
DB 132 AAAEAEQLL 140

RESULT 11
AAG43599
ID AAG43599 standard; protein; 245 AA.
AC AAG43599;
XX
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54513.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridization assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132662P.
PR 11-MAY-1999; 99US-0132663P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.

18-MAY-1999; 99US-0134768P.
19-MAY-1999; 99US-0134941P.
20-MAY-1999; 99US-0135124P.
21-MAY-1999; 99US-0135353P.
24-MAY-1999; 99US-0135629P.
25-MAY-1999; 99US-0136021P.
27-MAY-1999; 99US-0136392P.
28-MAY-1999; 99US-0136782P.
01-JUN-1999; 99US-0137222P.
03-JUN-1999; 99US-0137528P.
04-JUN-1999; 99US-0137502P.
07-JUN-1999; 99US-0137724P.
08-JUN-1999; 99US-0138094P.
10-JUN-1999; 99US-0138540P.
10-JUN-1999; 99US-0138847P.
14-JUN-1999; 99US-0139119P.
16-JUN-1999; 99US-0139452P.
16-JUN-1999; 99US-0139453P.
17-JUN-1999; 99US-0139492P.
18-JUN-1999; 99US-0139454P.
18-JUN-1999; 99US-0139455P.
18-JUN-1999; 99US-0139456P.
18-JUN-1999; 99US-0139457P.
18-JUN-1999; 99US-0139458P.
18-JUN-1999; 99US-0139459P.
18-JUN-1999; 99US-0139460P.
18-JUN-1999; 99US-0139461P.
18-JUN-1999; 99US-0139462P.
18-JUN-1999; 99US-0139463P.
18-JUN-1999; 99US-0139750P.
18-JUN-1999; 99US-0139763P.
21-JUN-1999; 99US-0139817P.
22-JUN-1999; 99US-0139899P.
23-JUN-1999; 99US-0140353P.
23-JUN-1999; 99US-0140354P.
24-JUN-1999; 99US-0140695P.
28-JUN-1999; 99US-0140823P.
29-JUN-1999; 99US-0140991P.
30-JUN-1999; 99US-0141287P.
01-JUL-1999; 99US-0141842P.
01-JUL-1999; 99US-0142154P.
02-JUL-1999; 99US-0142055P.
06-JUL-1999; 99US-0142390P.
08-JUL-1999; 99US-0142803P.
09-JUL-1999; 99US-0142920P.
12-JUL-1999; 99US-0142977P.
13-JUL-1999; 99US-0143542P.
14-JUL-1999; 99US-0143624P.
15-JUL-1999; 99US-0144005P.
16-JUL-1999; 99US-0144085P.
16-JUL-1999; 99US-0144086P.
19-JUL-1999; 99US-0144325P.
19-JUL-1999; 99US-0144331P.
19-JUL-1999; 99US-0144332P.
19-JUL-1999; 99US-0144333P.
19-JUL-1999; 99US-0144334P.
19-JUL-1999; 99US-0144335P.
20-JUL-1999; 99US-0144352P.
20-JUL-1999; 99US-0144632P.
20-JUL-1999; 99US-0144884P.
21-JUL-1999; 99US-0144814P.
21-JUL-1999; 99US-0145086P.
21-JUL-1999; 99US-0145088P.
22-JUL-1999; 99US-0145085P.
22-JUL-1999; 99US-0145087P.
22-JUL-1999; 99US-0145089P.
22-JUL-1999; 99US-0145192P.
23-JUL-1999; 99US-0145145P.
23-JUL-1999; 99US-0145218P.
23-JUL-1999; 99US-0145224P.
26-JUL-1999; 99US-0145276P.
27-JUL-1999; 99US-0145913P.
27-JUL-1999; 99US-0145918P.
```

```
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147203P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0160814P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160815P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.

PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 245;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAARAEQLL 937
Db 133 AAARAEQLL 141

RESULT 12
AAB92626
ID AAB92626 standard; protein; 279 AA.
AC AAB92626;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:10923.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 10923; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by
```

CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

XX
SQ Sequence 279 AA;

Query Match 0.9%; Score 9; DB 4; Length 279;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952
Db 18 SEKSSSES 26

RESULT 13

ADJ69129
ID ADJ69129 standard; protein; 279 AA.

XX
AC ADJ69129;

XX
DT 06-MAY-2004 (first entry)

XX
DE Human heat mitochondrial protein as a therapeutic target SeqID935.

XX
KW mitochondrial; human; screening assay; diabetes mellitus;
KW Huntington's disease; osteoarthritis;
KW Leber's hereditary optic neuropathy; LHON;
KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW osteopathic; ophthalmological; cytostatic.

XX
OS Homo sapiens.

XX
PN WO2003087768-A2.

XX
PD 23-OCT-2003.

XX
PF 04-APR-2003; 2003WO-US010870.

XX
PR 12-APR-2002; 2002US-0372843P.

XX
PR 17-JUN-2002; 2002US-0389987P.

XX
PR 20-SEP-2002; 2002US-0412418P.

XX
PA (MITO-) MITOKOR.

XX
PA (BUCK-) BUCK INST AGE RES.

XX
PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;

XX
PI Warnock DE;

XX
XX WPI; 2003-845369/78.

XX
XX

XX
XX Identifying a mitochondrial target for drug screening assays and for

XX
XX treating diseases associated with altered mitochondrial function,

XX
XX comprises detecting a modified polypeptide in a sample and correlating

XX
XX with the disease.

XX
XX Claim 1; SEQ ID NO 935; 180pp; English.

XX
XX This invention relates to novel mitochondrial targets that can be used
XX for therapeutic intervention in treating a disease associated with
XX altered mitochondrial function. Specifically, it refers to a method for
XX identifying proteins of the human heart mitochondrial proteome that are
XX useful for drug screening assays, as well as therapeutic targets. The
XX present invention describes a method for identifying such proteins that
XX can be used in the treatment of various diseases associated with altered
XX mitochondrial function including diabetes mellitus, Huntington's disease,
XX osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
XX encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy

CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, nootropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC cytostatic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.

XX
SQ Sequence 279 AA;

Query Match 0.9%; Score 9; DB 7; Length 279;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952
Db 18 SEKSSSES 26

RESULT 14

AAG44240

ID AAG44240 standard; protein; 340 AA.

XX
AC AAG44240;

XX
DT 18-OCT-2000 (first entry)

XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55391.

XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-00301439.

XX
PR 25-FEB-1999; 99US-0121825P.

XX
PR 05-MAR-1999; 99US-0123180P.

XX
PR 09-MAR-1999; 99US-0123548P.

XX
PR 23-MAR-1999; 99US-0125788P.

XX
PR 25-MAR-1999; 99US-0126284P.

XX
PR 29-MAR-1999; 99US-0126785P.

XX
PR 01-APR-1999; 99US-0127462P.

XX
PR 06-APR-1999; 99US-0128234P.

XX
PR 08-APR-1999; 99US-0128714P.

XX
PR 16-APR-1999; 99US-0129845P.

XX
PR 19-APR-1999; 99US-0130077P.

XX
PR 21-APR-1999; 99US-0130449P.

XX
PR 23-APR-1999; 99US-0130510P.

XX
PR 28-APR-1999; 99US-0130891P.

XX
PR 30-APR-1999; 99US-0131449P.

XX
PR 30-APR-1999; 99US-0132048P.

XX
PR 04-MAY-1999; 99US-0132407P.

XX
PR 05-MAY-1999; 99US-0132484P.

XX
PR 06-MAY-1999; 99US-0132486P.

XX
PR 06-MAY-1999; 99US-0132487P.

XX
PR 07-MAY-1999; 99US-0132863P.

XX
PR 11-MAY-1999; 99US-0134256P.

XX
PR 14-MAY-1999; 99US-0134218P.

XX
PR 14-MAY-1999; 99US-0134219P.

XX
PR 14-MAY-1999; 99US-0134221P.

XX
PR 14-MAY-1999; 99US-0134370P.

XX
PR 18-MAY-1999; 99US-0134768P.

XX
PR 19-MAY-1999; 99US-0134941P.

XX
PR 20-MAY-1999; 99US-0135124P.

XX
PR 21-MAY-1999; 99US-0135353P.

XX
PR 24-MAY-1999; 99US-0135629P.

XX
PR 25-MAY-1999; 99US-0136021P.

XX
PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140595P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147418P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159291P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

```
Query Match          0.9%; Score 9; DB 3; Length 340;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAARAEQLL 937
Db 123 AAARAEQLL 131

RESULT 15
AAG44239
ID AAG44239 standard; protein; 349 AA.
AC AAG44239;
XX
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55390.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 07-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139452P.
XX 16-JUN-1999; 99US-0139453P.
XX 17-JUN-1999; 99US-0139492P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.
XX 18-JUN-1999; 99US-0139456P.
XX 18-JUN-1999; 99US-0139457P.
XX 18-JUN-1999; 99US-0139458P.
XX 18-JUN-1999; 99US-0139459P.
XX 18-JUN-1999; 99US-0139460P.
XX 18-JUN-1999; 99US-0139461P.
XX 18-JUN-1999; 99US-0139462P.
XX 18-JUN-1999; 99US-0139463P.
XX 18-JUN-1999; 99US-0139750P.
XX 21-JUN-1999; 99US-0139763P.
XX 22-JUN-1999; 99US-0139817P.
XX 23-JUN-1999; 99US-0139899P.
XX 23-JUN-1999; 99US-0140353P.
XX 23-JUN-1999; 99US-0140354P.
XX 24-JUN-1999; 99US-0140695P.
XX 28-JUN-1999; 99US-0140823P.
XX 29-JUN-1999; 99US-0140991P.
XX 30-JUN-1999; 99US-0141287P.
XX 01-JUL-1999; 99US-0141842P.
XX 01-JUL-1999; 99US-0142154P.
XX 02-JUL-1999; 99US-0142055P.
XX 06-JUL-1999; 99US-0142390P.
XX 08-JUL-1999; 99US-0142803P.
XX 09-JUL-1999; 99US-0142920P.
XX 12-JUL-1999; 99US-0142977P.
XX 13-JUL-1999; 99US-0143542P.
XX 14-JUL-1999; 99US-0143624P.
XX 15-JUL-1999; 99US-0144005P.
XX 16-JUL-1999; 99US-0144085P.
XX 16-JUL-1999; 99US-0144086P.
XX 19-JUL-1999; 99US-0144325P.
XX 19-JUL-1999; 99US-0144331P.
XX 19-JUL-1999; 99US-0144332P.
XX 19-JUL-1999; 99US-0144333P.
XX 19-JUL-1999; 99US-0144334P.
XX 19-JUL-1999; 99US-0144335P.
XX 20-JUL-1999; 99US-0144352P.
XX 20-JUL-1999; 99US-0144632P.
XX 20-JUL-1999; 99US-0144633P.
XX 21-JUL-1999; 99US-0144814P.
XX 21-JUL-1999; 99US-0145086P.
XX 21-JUL-1999; 99US-0145088P.
XX 22-JUL-1999; 99US-0145087P.
XX 22-JUL-1999; 99US-0145089P.
XX 22-JUL-1999; 99US-0145192P.
XX 23-JUL-1999; 99US-0145145P.
XX 23-JUL-1999; 99US-0145218P.
XX 23-JUL-1999; 99US-0145224P.
XX 26-JUL-1999; 99US-0145276P.
XX 27-JUL-1999; 99US-0145913P.
XX 27-JUL-1999; 99US-0145918P.
XX 27-JUL-1999; 99US-0145919P.
XX 28-JUL-1999; 99US-0145951P.
XX 02-AUG-1999; 99US-0146386P.
XX 02-AUG-1999; 99US-0146388P.
XX 02-AUG-1999; 99US-0146389P.
XX 03-AUG-1999; 99US-0147038P.
XX 04-AUG-1999; 99US-0147204P.
XX 04-AUG-1999; 99US-0147302P.
XX 05-AUG-1999; 99US-0147322P.
XX 05-AUG-1999; 99US-0147260P.
XX 06-AUG-1999; 99US-0147303P.
XX 06-AUG-1999; 99US-0147416P.
XX 09-AUG-1999; 99US-0147493P.
XX 09-AUG-1999; 99US-0147935P.
```

```
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 20-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151338P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAGAEQQL 937
|||||||

Db 132 AAAGAEQQL 140

RESULT 16
AAW40386
ID AAM40386 standard; protein; 361 AA.
XX
AC AAM40386;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3531.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
WPI; 2001-442253/47.
DR N-PSDB; AAI59542.
XX.
PT Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.
XX
PS Example 6; SEQ ID NO 3531; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
encoded polypeptides (AAM38642-AAW42213) with nootropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: immune system suppression,
Activin/inhibin activity; chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemias and
C.N.S disorders. Note: The sequence data for this patent did not form
part of the printed specification
XX
SQ Sequence 361 AA;

Query Match 0.9%; Score 9; DB 4; Length 361;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      944 SEKSSSSSES 952
Db      18 SEKSSSSSES 26
      |||||
RESULT 17
AAG44241
ID AAG444241 standard; protein; 370 AA.
XX      AC
XX      AAG44241;
XX      18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55392.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 200DEP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131448P.
PR 30-APR-1999; 99US-0134048P.
PR 30-APR-1999; 99US-0134407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142330P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147935P.
PR 09-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
```


PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151338P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154778P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0158293P.
 PR 13-OCT-1999; 99US-0158294P.
 PR 13-OCT-1999; 99US-0158295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 370;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAAEAEQLL 937
 Db 153 AAAAEAEQLL 161

RESULT 18

AAB48789
 ID AAB48789 standard; protein; 557 AA.
 XX
 AC AAB48789;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human prostate cancer-predisposing protein, CA7 CG04.
 XX
 KW Human; CA7 CG04 gene; prostate cancer-predisposing gene;
 KW chromosome 1 HPC1 region; inherited early onset prostate cancer;
 KW guanosine exchange factor protein; GEF homologue; ras activator;
 KW diagnosis; anticancer drug screening; mutation screening.
 XX
 OS Homo sapiens.
 XX
 FN WO200069879-A2.
 XX
 PD 23-NOV-2000.
 XX
 PF 11-MAY-2000; 2000WO-US012917.
 XX
 PR 14-MAY-1999; 99US-0134209P.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
 XX
 PI Tavtigian SV, Swedlund B, Simard J, Rommens JM;
 DR WPI; 2001-016208/02.
 DR N-PSDB; AAC87385.
 XX
 PT Novel human prostate cancer predisposing gene useful for diagnosis and
 PT prognosis of cancer, especially, prostate cancer and for screening drugs
 PT for cancer therapy.
 XX
 PS Claim 1; Page 91-92; 96pp; English.
 CC
 CC The invention relates to a human prostate cancer-predisposing gene,
 CC designated CA7 CG04 (cDNA given in AAC87385), and the CA7 CG04 protein
 CC (AAB48789). The CA7 CG04 protein has homology with guanosine exchange
 CC factor (GEF) proteins, and is therefore thought to function as an
 CC activator of ras. The CA7 CG04 gene is located on chromosome 1 in the
 CC HPC1 region, which is linked with inherited early onset prostate cancer.
 CC The invention also relates to exons 1-19 of the CA7 CG04 gene (AAC87386-
 CC C87404), allelic variants and mutants of the CA7 CG04 gene, expression
 CC vectors and host cells comprising a CA7 CG04 DNA, recombinant production
 CC of the CA7 CG04 protein, anti-CA7 CG04 antibodies, CA7 CG04 primers and
 CC probes, and methods of screening for potential anticancer drugs which can
 CC inhibit the ability of a mutant CA7 CG04 to activate ras. The CA7 CG04
 CC cDNA and protein are useful for screening potential anticancer drugs and
 CC for screening the CA7 CG04 gene for mutations which are useful in
 CC diagnosing a predisposition to cancer, especially prostate cancer. The
 CC present sequence represents human CA7 CG04 protein
 XX
 SQ Sequence 557 AA;

Query Match 0.9%; Score 9; DB 4; Length 557;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952
 Db 18 SEKSSSES 26

RESULT 19
 ABB97502
 ID ABB97502 standard; protein; 557 AA.
 XX
 AC ABB97502;
 XX

DT 27-JUN-2002 (first entry)

XX Novel human protein SEQ ID NO: 770.
XX
XX Human; antianemic; vulnerary; antiinflammatory; immunomodulator;
KW antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.
XX
OS Homo sapiens.
XX
XX WO200222660-A2.
XX
XX PD
XX 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-0026015.
XX
XX 11-SEP-2000; 2000US-00659671.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
DR N-PSDB; ABN32688.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
XX Claim 20; SEQ ID NO 770; 509pp; English.
XX
XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention
XX
SQ Sequence 557 AA;

Query Match 0.9%; Score 9; DB 5; Length 557;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952
Db 18 SEKSSSES 26
|||||

RESULT 20
AAG49069
ID AAG49069 standard; protein; 581 AA.
XX
AC AAG49069;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 62038.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX

PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137548P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 28-JUN-1999; 99US-0140695P.
PR 29-JUN-1999; 99US-0140823P.
PR 30-JUN-1999; 99US-0140991P.
PR 01-JUL-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.

```
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149168P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.

PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 581;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAREAQQLL 937
Db 123 AAAREAQQLL 131

RESULT 21
ADJ50969
ID ADJ50969 standard; protein; 583 AA.
AC ADJ50969;
XX
XX 06-MAY-2004 (first entry)
DE Human novel protein NOV12a.
XX
XX Human; NOVX; autoimmune disease; Alzheimer's disease; stroke; allergy;
XX Parkinson's disease; Huntington's disease; multiple sclerosis; anxiety;
XX pain; diabetes; graft versus host disease; pancreatitis; obesity; ulcer;
XX anaemia; cancer; viral infection; bacterial infection;
XX parasitic infection.
OS Homo sapiens.
XX
XX US2004030096-A1.
XX
XX 12-FEB-2004.
XX
XX 01-AUG-2002; 2002US-00210281.
XX
XX 02-AUG-2001; 2001US-0309501P.
XX 03-AUG-2001; 2001US-0310291P.
XX 08-AUG-2001; 2001US-0310951P.
```

09-AUG-2001; 2001US-0311292P.
PR 13-AUG-2001; 2001US-0311979P.
PR 14-AUG-2001; 2001US-0311203P.
PR 17-AUG-2001; 2001US-03113201P.
PR 20-AUG-2001; 2001US-03113643P.
PR 20-AUG-2001; 2001US-0313703P.
PR 21-AUG-2001; 2001US-0314031P.
PR 23-AUG-2001; 2001US-0314466P.
PR 28-AUG-2001; 2001US-0315403P.
PR 29-AUG-2001; 2001US-0315853P.
PR 05-MAR-2002; 2002US-0361775P.
PR 05-MAR-2002; 2002US-0361832P.
XX (GORM/) GORMAN L.
PA (ZERRH/) ZERHUSEN B D.
PA (EDIN/) EDINGER S R.
PA (PADI/) PADIGARU M.
PA (GUOX/) GUO X.
PA (KEKU/) KEKUDA R.
PA (ZHON/) ZHONG M.
PA (PATI/) PATTURAJAN M.
PA (MILL/) MILLER C E.
PA (JIWW/) JI W.
PA (PENA/) PENNA C E A.
PA (BURG/) BURGESS C E.
PA (SCIO/) SCIORE P.
PA (STON/) STONE D J.
PA (TAUF/) TAUFIER R J.
PA (CASM/) CASMAN S J.
PA (ROTH/) ROTHENBERG M E.
PA (MALY/) MALYANKAR U M.
PA (BOLD/) BOLDOG F L.
XX
PI Gorman L, Zerhusen BD, Edinger SR, Padigaru M, Guo X, Kekuda R;
PI Zhong M, Patturajan M, Miller CE, Ji W, Pena CE, Burgess CE;
PI Sciore P, Stone DJ, Taupier RJ, Casman SJ, Rothenberg ME;
PI Malyankar UM, Boldog FL;
XX WPI, 2004-168942/16.
DR N-PSDB; ADJ50968.
XX
New NOVX polypeptides and polynucleotides, useful in diagnosing, treating
PT or preventing diseases or conditions, e.g. autoimmune disease,
PT Alzheimer's disease, diabetes, graft versus host disease, cancer or viral
PT or bacterial infections.
XX
Claim 2; SEQ ID NO 34; 342pp; English.
XX
The invention relates to an isolated NOVX polypeptide (of 44 disclosed)
CC comprising its mature form, a sequence having at least 95% sequence
CC identity to NOVX or a sequence comprising one or more conservative
CC substitutions in the amino acid sequence of NOVX. Also included are a
CC composition comprising NOVX and a carrier, a kit comprising, in one or
CC more containers, the composition, a method of identifying an agent that
CC binds to NOVX, a method for identifying a potential therapeutic agent for
CC use in treatment of a pathology related to aberrant expression or
CC aberrant physiological interactions of NOVX, a method for screening for a
CC modulator of activity of or of latency or predisposition to a pathology
CC associated with NOVX, a method for modulating the activity of NOVX, a
CC method of treating or preventing a pathology associated with NOVX or a
CC pathological state in a mammal, an isolated nucleic acid molecule
CC encoding a NOVX protein, a vector comprising the nucleic acid molecule,
CC a cell comprising the vector, an antibody that immunospecifically binds
CC to NOVX, a method for determining the presence or amount of NOVX or the
CC nucleic acid molecule in a sample, a method for determining the presence
CC of or predisposition to a disease associated with altered levels of
CC expression of NOVX or the nucleic acid molecule in a first mammalian
CC subject and a method of producing NOVX (comprising culturing the cell
CC under conditions that lead to expression of the polypeptide). NOVX is
CC useful in the manufacture of a medicament for treating a syndrome
CC associated with a human disease associated with NOVX. The polypeptides
CC and nucleic acid molecules are useful in diagnosing, treating or
CC preventing diseases or conditions, e.g. autoimmune disease, Alzheimer's

CC disease, stroke, allergies, Parkinson's disease, Huntington's disease,
CC multiple sclerosis, anxiety, pain, diabetes, graft versus host disease,
CC pancreatitis, obesity, ulcers, anaemia, cancer, viral or bacterial and
CC parasitic infections (many more diseases and disorders are listed in the
CC specification). The present sequence represents a NOVX protein.
XX
SQ Sequence 583 AA;

Query Match 0.9%; Score 9; DB 8; Length 583;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952
Db 18 SEKSSSES 26
|||||

RESULT 22

AAG49068
ID AAG49068 standard; protein; 590 AA.

XX
AC AAG49068;

XX
DT 18-OCT-2000 (first entry)

XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62037.

XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX
FN EP1033405-A2.

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-00301439.

XX
PR 25-FEB-1999; 99US-0121825P.

XX
PR 05-MAR-1999; 99US-0123180P.

XX
PR 09-MAR-1999; 99US-0123548P.

XX
PR 23-MAR-1999; 99US-0125788P.

XX
PR 25-MAR-1999; 99US-0126284P.

XX
PR 29-MAR-1999; 99US-0126785P.

XX
PR 01-APR-1999; 99US-0127462P.

XX
PR 06-APR-1999; 99US-0128234P.

XX
PR 08-APR-1999; 99US-0128714P.

XX
PR 16-APR-1999; 99US-0129845P.

XX
PR 19-APR-1999; 99US-0130077P.

XX
PR 21-APR-1999; 99US-0130449P.

XX
PR 23-APR-1999; 99US-0130510P.

XX
PR 28-APR-1999; 99US-0130891P.

XX
PR 30-APR-1999; 99US-0131449P.

XX
PR 30-APR-1999; 99US-0132048P.

XX
PR 04-MAY-1999; 99US-0132407P.

XX
PR 06-MAY-1999; 99US-0132484P.

XX
PR 05-MAY-1999; 99US-0132485P.

XX
PR 06-MAY-1999; 99US-0132486P.

XX
PR 07-MAY-1999; 99US-0132487P.

XX
PR 11-MAY-1999; 99US-0132863P.

XX
PR 14-MAY-1999; 99US-0134218P.

XX
PR 14-MAY-1999; 99US-0134219P.

XX
PR 14-MAY-1999; 99US-0134221P.

XX
PR 14-MAY-1999; 99US-0134370P.

XX
PR 18-MAY-1999; 99US-0134768P.

XX
PR 19-MAY-1999; 99US-0134941P.

XX
PR 20-MAY-1999; 99US-0135124P.

XX
PR 21-MAY-1999; 99US-0135353P.

XX
PR 24-MAY-1999; 99US-0135629P.

XX
PR 25-MAY-1999; 99US-0136021P.

XX
PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

```
Query Match          0.9%; Score 9; DB 3; Length 590;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAREQLL 937
DB 132 AAAREQLL 140
|||||

RESULT 23
ABB93671
ID ABB93671 standard; protein; 590 AA.
XX
AC ABB93671;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2882.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP009892.
XX
PR 28-AUG-2001; 2001WO-EP009892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidner M;
DR WPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.
XX
PS Claim 5; SEQ ID NO 2882; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides
XX
SQ Sequence 590 AA;

Query Match          0.9%; Score 9; DB 5; Length 590;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAREQLL 937
DB 132 AAAREQLL 140
|||||

RESULT 24
ABM85378
ID ABM85378 standard; protein; 608 AA.
XX
AC ABM85378;
XX
DT 18-NOV-2004 (first entry)
XX
```

```
XX Mouse protein sequence mCP21067.
XX DE
XX Cytostatic; carcinoma; lymphoma; cancer; murine.
XX KW
XX Mus musculus.
XX OS
XX WO2003073826-A2.
XX PN
XX PD 12-SEP-2003.
XX
XX 28-FEB-2003; 2003WO-US006235.
XX PF
XX 01-MAR-2002; 2002US-00087192.
XX PR
XX (SAGR-) SAGRES DISCOVERY.
XX PA
XX Morris DW;
XX PI
XX WPI; 2003-328604/31.
XX DR
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX PT comprises a nucleotide sequence.
XX
XX Claim 5; SEQ ID NO 549; Opp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CAP. Note:
CC This patent is an equivalent to basic patent US2002182586A1, for which no
CC sequence data was published
XX
SQ Sequence 608 AA;

Query Match          0.9%; Score 9; DB 7; Length 608;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSSES 952
DB 36 SEKSSSSSES 44
|||||

RESULT 25
AAG49067
ID AAG49067 standard; protein; 611 AA.
XX
AC AAG49067;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62036.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX
XX Arabidopsis thaliana.
XX OS
XX EP1033405-A2.
XX PN
XX PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX PF
```

XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.

```
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158233P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161403P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 611;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAAEQQL 937
DB 153 AAAAEQQL 161

RESULT 26
AAW93890
ID AAW93890 standard; protein; 692 AA.
AC AAW93890;
XX
XX
DT 25-JUN-1999 (first entry)
DE Human HG38 protein.
KW HG38; human; G-protein coupled glycoprotein hormone receptor; brain;
KW endocrine system; skeletal muscle; placenta; development;
KW receptor activity modulator.
OS Homo sapiens.
XX
XX WO9915660-A1.
XX
PD 01-APR-1999.
XX
XX PF 24-SEP-1998; 98WO-US019979.
XX
XX PR 24-SEP-1997; 97US-0059863P.
XX
XX (MERI ) MERCK & CO INC.
XX
XX PI Liu Q, Bailey WJ, McDonald TP;
XX

Query Match 0.9%; Score 9; DB 2; Length 692;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
DB 174 LTEIPVQAF 182

RESULT 27
ADB80464
ID ADB80464 standard; protein; 883 AA.
XX
XX ADB80464;
AC ADB80464;
XX
XX DT 04-DEC-2003 (first entry)
XX
XX DE Ovarian cancer-associated protein #14.
XX
XX KW cytostatic; gene therapy; vaccine; ovarian cancer; diagnosis;
XX post-operative chemotherapy; radiation therapy; tumour prognosis;
XX pre-cancerous lesion detection.
OS Homo sapiens.
XX
XX PN WO2002102235-A2.
XX
XX PD 27-DEC-2002.
XX
XX PF 18-JUN-2002; 2002WO-US019297.
XX
XX PR 18-JUN-2001; 2001US-0299234P.
XX
XX PR 27-AUG-2001; 2001US-0315287P.
XX
XX PR 05-SEP-2001; 2001US-0317544P.
XX
XX PR 13-NOV-2001; 2001US-0350666P.
XX
XX PR 12-APR-2002; 2002US-0372246P.
XX
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX
XX PI Mack DH, Gish KC;
XX
XX DR N-PSDB; ADB80463.
XX
XX PT Detecting an ovarian cancer-associated transcript in a cell from a
XX patient, comprises contacting a biological sample from the patient with a
XX polynucleotide that hybridizes to an ovarian cancer gene.
XX
XX PS Claim 13; Page 284; 332pp; English.
XX
XX The invention relates to a method of detecting an ovarian cancer-
XX associated transcript in a cell from a patient, by contacting a
XX biological sample from the patient with a polynucleotide that selectively
XX hybridizes to a sequence at least 80% identical to any of one of 80
XX nucleic acid sequences given in the specification. The method is useful
```


CC in diagnosing ovarian cancer and in identifying and using agents and/or
 CC targets that inhibit ovarian cancer. The nucleic acid molecule,
 CC polypeptide and the antibody may also be used in detecting ovarian
 CC cancers, monitoring and early detection of relapse following treatment,
 CC monitoring response to therapy, selecting patients for post-operative
 CC chemotherapy or radiation therapy, in selecting mode of therapy,
 CC determining tumour prognosis, early detection of pre-cancerous lesions,
 CC and as vaccines. This sequence corresponds to one of the proteins used
 CC for the detection method of the invention.

XX Sequence 883 AA;

Query Match 0.9%; Score 9; DB 7; Length 883;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
 |||||
 Db 174 LTEIPVQAF 182

RESULT 28

ADN40012
 ID ADN40012 standard; protein; 883 AA.

AC ADN40012;

DT 17-JUN-2004 (first entry)

DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO: C382.

XX Human; differential expression; cancer; angiogenic disorder;
 KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;
 KW inflammatory disease; autoimmune disease;
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;
 KW detection; diagnosis; prognosis; drug screening; drug targeting;
 KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;
 KW vulnery; gene therapy; vaccine.

XX Homo sapiens.

XX WO2003042661-A2.

XX 22-MAY-2003.

XX 13-NOV-2002; 2002WO-US036810.

XX 13-NOV-2001; 2001US-0350666P.
 PR 21-NOV-2001; 2001US-0332464P.
 PR 29-NOV-2001; 2001US-0334393P.
 PR 03-DEC-2001; 2001US-0335394P.
 PR 14-DEC-2001; 2001US-0340376P.
 PR 08-JAN-2002; 2002US-0347211P.
 PR 10-JAN-2002; 2002US-0347349P.
 PR 13-FEB-2002; 2002US-0355250P.
 PR 08-FEB-2002; 2002US-0356714P.
 PR 20-FEB-2002; 2002US-0359077P.
 PR 29-MAR-2002; 2002US-0368809P.
 PR 04-APR-2002; 2002US-0370110P.
 PR 12-APR-2002; 2002US-0372246P.
 PR 05-JUN-2002; 2002US-0386614P.
 PR 16-JUL-2002; 2002US-0396839P.
 PR 22-JUL-2002; 2002US-0397775P.
 PR 22-JUL-2002; 2002US-0397845P.
 PR 09-SEP-2002; 2002US-0409450P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;
 PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;

XX WPI; 2003-468649/44.

DR N-PSDB; ADN39795.

XX

PT Determining the presence or absence of a pathological cell in a patient,
 PT useful for diagnosing, prognosing or treating cancer, comprises detecting
 PT a nucleic acid in a biological sample.

XX Claim 12; SEQ ID NO C382; 1385pp; English.

XX The invention relates to nucleic acids and proteins (ADN3863-ADN40064)
 CC whose expression is upregulated or downregulated in specific cancers or
 CC other diseases such as angiogenic or fibrotic disorders, and to methods
 CC of determining the presence or absence of a pathological cell in a
 CC patient by detecting a nucleic acid at least 80% identical to those of
 CC the invention or by detecting a polypeptide of the invention. The
 CC invention also relates to expression vectors and host cells comprising a
 CC nucleic acid of the invention; antibodies which specifically bind a
 CC polypeptide of the invention; use of such antibodies for drug targeting;
 CC and methods of screening for modulators of activity or expression of the
 CC polypeptides and nucleic acids. The nucleic acids, polypeptides,
 CC antibodies and methods are useful for diagnosing, prognosing and treating
 CC cancer and other conditions such as psoriasis, ischaemia, heart disease,
 CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal
 CC neovascularisation syndromes, scarring and uterine fibroids. They may
 CC also be useful in wound healing and in contraception. The present
 CC sequence represents a polypeptide of the invention.

XX Sequence 883 AA;

Query Match 0.9%; Score 9; DB 7; Length 883;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
 |||||
 Db 174 LTEIPVQAF 182

RESULT 29

ADN39166
 ID ADN39166 standard; protein; 883 AA.

AC ADN39166;

XX 17-JUN-2004 (first entry)

XX Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO: 484.

XX Human; differential expression; cancer; angiogenic disorder;
 KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;
 KW inflammatory disease; autoimmune disease;
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;
 KW detection; diagnosis; prognosis; drug screening; drug targeting;
 KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;
 KW vulnery; gene therapy; vaccine.

XX Homo sapiens.

XX WO2003042661-A2.

XX 22-MAY-2003.

XX 13-NOV-2002; 2002WO-US036810.

XX 13-NOV-2001; 2001US-0350666P.
 PR 21-NOV-2001; 2001US-0332464P.
 PR 29-NOV-2001; 2001US-0334393P.
 PR 03-DEC-2001; 2001US-0335394P.
 PR 14-DEC-2001; 2001US-0340376P.
 PR 08-JAN-2002; 2002US-0347211P.
 PR 10-JAN-2002; 2002US-0347349P.
 PR 08-FEB-2002; 2002US-0355250P.
 PR 13-FEB-2002; 2002US-0356714P.
 PR 20-FEB-2002; 2002US-0359077P.
 PR 29-MAR-2002; 2002US-0368809P.

CC and AAA30775-A30779). The mutant proteins of the invention contain a
 CC mutation in a portion of the protein comprising intracellular loop 3
 CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
 CC is substituted for an endogenous residue in IC3 at a position 16 amino
 CC acids N-terminal of an endogenous proline in TM6 to form a sequence X-
 CC (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or
 CC Ala, and is preferably Lys. When the endogenous residue at this position
 CC is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
 CC amino acid stretch between the substituted amino acid and the Pro may be
 CC endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
 CC residues. The constitutively active GPCRs are useful for identifying
 CC antagonists, agonists and partial agonists for use as pharmaceutical
 CC agents. The mutant proteins are also useful in research settings for
 CC elucidating the roles of the receptors in normal and diseased conditions.
 CC Antagonists for a particular GPCR are useful for treating diseases and
 CC disorders associated with that receptor. Because the novel mutant GPCRs
 CC are constitutively active, they can be used directly for screening of
 CC compounds without the need for endogenous ligands. The present sequence
 CC represents a human wild-type GPCR referred to in an exemplification of
 CC the invention

XX Sequence 907 AA;

Query Match 0.9%; Score 9; DB 3; Length 907;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
 DB 174 LTEIPVQAF 182
 |||||

RESULT 32

AAAY90687
 ID AAAY90687 standard; protein; 907 AA.

XX AC AAAY90687;

DT 21-AUG-2000 (first entry)

DE Human mutant G protein-coupled receptor HG38 (V765K).

XX G protein-coupled receptor; GPCR; constitutively active;
 KW intracellular loop 3; transmembrane domain 6; drug screening; agonist;
 KW antagonist; mutant; mutein.

OS Homo sapiens.
 OS Synthetic.

XX WO200022129-A1.

XX 20-APR-2000.

XX 12-OCT-1999; 99WO-US023939.

XX 13-OCT-1998; 98US-00170496.

XX (AREN-) ARENA PHARM INC.

XX Behan DP, Chalmers DT, Liaw CW;

XX WPI; 2000-329165/28.

XX N-PSDB; AAA30779.

XX Non-endogenous constitutively activated human G protein-coupled
 PT receptors, useful for identifying agonists for use as pharmaceutical
 PT agents.

XX Example 2; Page 332-335; 341pp; English.

XX The invention relates to constitutively active, non-endogenous versions
 CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAAY90643-
 CC AAAY90677 and AAAY90683-Y90687), and to DNA encoding them (AAA30709-A30743

CC and AAA30775-A30779). The mutant proteins of the invention contain a
 CC mutation in a portion of the protein comprising intracellular loop 3
 CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
 CC is substituted for an endogenous residue in IC3 at a position 16 amino
 CC acids N-terminal of an endogenous proline in TM6 to form a sequence X-
 CC (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or
 CC Ala, and is preferably Lys. When the endogenous residue at this position
 CC is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
 CC amino acid stretch between the substituted amino acid and the Pro may be
 CC endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
 CC residues. The constitutively active GPCRs are useful for identifying
 CC antagonists, agonists and partial agonists for use as pharmaceutical
 CC agents. The mutant proteins are also useful in research settings for
 CC elucidating the roles of the receptors in normal and diseased conditions.
 CC Antagonists for a particular GPCR are useful for treating diseases and
 CC disorders associated with that receptor. Because the novel mutant GPCRs
 CC are constitutively active, they can be used directly for screening of
 CC compounds without the need for endogenous ligands. Sequences AAAY90643-
 CC AAAY90677 and AAAY90683-Y90687 the mutant human GPCRs of the invention
 XX

Sequence 907 AA;

Query Match 0.9%; Score 9; DB 3; Length 907;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
 DB 174 LTEIPVQAF 182
 |||||

RESULT 33

ABP81968

ID ABP81968 standard; protein; 907 AA.

XX AC ABP81968;

DT 04-MAR-2003 (first entry)

DE Human G protein-coupled receptor GPR49 protein SEQ ID NO:422.

XX G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
 KW G protein-coupled receptor modulator; antibody; immune-related disease;
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;
 KW immunological-related cell proliferative disease; autoimmune disease;
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
 KW ulcer.

XX Homo sapiens.

XX WO200261087-A2.

XX 08-AUG-2002.

XX 19-DEC-2001; 2001WO-US050107.

XX 19-DEC-2000; 2000US-0257144P.

XX (LIFE-) LIFESPAN BIOSCIENCES INC.

XX Burmer GC, Roush CL, Brown JP;

XX WPI; 2003-046718/04.

XX N-PSDB; ABZ42816.

XX New isolated antigenic peptides e.g., for G protein-coupled receptors
 PT (GPCR), useful for diagnosing and designing drugs for treating conditions
 PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or

PT autoimmune diseases.

XX Disclosure; Fig 1; 523pp; English.

XX

CC The present invention describes antigenic peptides (I) comprising: (a)

CC any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino

CC acids. Also described: (1) an assay for the detection of a particular G

CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;

CC and (2) an isolated antibody having high specificity and high affinity or

CC avidity for a particular GPCR. (I) can be used as GPCR modulators and in

CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an

CC antibody against a particular GPCR, and in the production of specific

CC antibodies. The peptides and antibodies are also useful for detecting the

CC presence or absence of corresponding GPCRs. The antigenic peptides for

CC GPCRs and antibodies are useful for diagnosing and designing drugs for

CC treating immune-related diseases, growth-related diseases, cell

CC regeneration-related disease, immunological-related cell proliferative

CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,

CC atherosclerosis, bacterial, fungal, protozoan or viral infections,

CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute

CC inflammation, allergies, Crohn's disease, diabetes, graft versus host

CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,

CC anxiety, depression, schizophrenia, dementia, mental retardation, memory

CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,

CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or

CC any other disorder in which GPCRs are involved. The antibodies may be

CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode

CC GPCR proteins given in ABP81675 to ABP82018, which are used in the

CC exemplification of the present invention

XX

SQ Sequence 907 AA;

Query Match 0.9%; Score 9; DB 6; Length 907;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600

DB 174 LTEIPVQAF 182

|||||

RESULT 34

ABO06467

ID ABO06467 standard; protein; 907 AA.

XX

AC ABO06467;

XX

DT 13-AUG-2003 (first entry)

XX

DE Human G-protein coupled receptor HG38.

XX

KW G-protein coupled receptor; GPCR; caudate nucleus related disorder;

XX neoplastic disorder; leukaemia; breast cancer; immune disorder; AIDS;

KW rheumatoid arthritis; neuronal disorder; Alzheimer's disease; receptor;

XX Parkinson's disease; respiratory disorder; bronchopulmonary disease;

KW pancreatic syndrome; ovarian disorder; ovarian carcinoma; colon disease;

XX Meig's syndrome; breast disease; thalamus related disorder; gene therapy;

KW amygdala related disorder; corpus callosum related disorder;

XX hippocampus related disorder.

XX

OS Homo sapiens.

XX

PN US2003027323-A1.

XX

PD 06-FEB-2003.

XX

PF 26-SEP-2001; 2001US-00965536.

XX

PR 27-SEP-2000; 2000US-0235832P.

XX

PR 16-JAN-2001; 2001US-0261781P.

XX

PR 19-JUL-2001; 2001US-030605P.

XX

PR 03-AUG-2001; 2001US-0310436P.

XX

PA (FEDE/) PEDER J N.

PA (MINT/) MINTIER G.

PA (RAWA/) RAMANATHAN C S.

PA (HAWK/) HAWKEN D R.

XX

Feder JN, Mintier G, Ramanathan CS, Hawken DR;

XX

WPI; 2003-466147/44.

XX

New G-protein coupled receptor polypeptides, designated as HGPRBMY5,

PT useful for preventing, treating or ameliorating a medical condition

PT related to the colon, breast, ovaries or immune system.

XX

Disclosure; Fig 8; 93pp; English.

XX

The invention relates to an isolated HGPRBMY5 polypeptide. The

CC polypeptides, polynucleotides and methods are useful for preventing,

CC treating or ameliorating a medical condition such as a neoplastic

CC disorder e.g. leukaemia and breast cancer; immune disorder e.g. AIDS and

CC rheumatoid arthritis; neuronal disorder e.g. Alzheimer's disease and

CC Parkinson's disease; respiratory disorder e.g. bronchopulmonary disease

CC and pancreatic syndrome; ovarian disorder e.g. ovarian carcinoma and

CC Meig's syndrome; colon disease; breast disease; thalamus related disorder

CC ; amygdala related disorder; corpus callosum related disorder; caudate

CC nucleus related disorder; hippocampus related disorder by administering

CC the GPCR polypeptide or its homologue. The present sequence represents

CC the amino acid sequence of a G-protein coupled receptor used to show

CC homology with the human G-protein coupled receptor, HGPRBMY

XX

SQ Sequence 907 AA;

Query Match 0.9%; Score 9; DB 6; Length 907;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600

DB 174 LTEIPVQAF 182

|||||

RESULT 35

ADC22797

ID ADC22797 standard; protein; 907 AA.

XX

AC ADC22797;

XX

DT 18-DEC-2003 (first entry)

XX

DE Human G protein-coupled receptor (GPCR) polypeptide #79.

XX

KW Human; G protein-coupled receptor; GPCR; transmembrane-6 region; TM6;

XX intracellular-3 region; IC3; receptor.

XX

OS Homo sapiens.

XX

PN US6555339-B1.

XX

PD 29-APR-2003.

XX

PF 13-OCT-1998; 98US-00170496.

XX

PR 14-APR-1997; 97US-00839449.

XX

PR 14-APR-1998; 98US-00060188.

XX

PR 26-JUN-1998; 98US-0090783P.

XX

PR 07-AUG-1998; 98US-0095677P.

XX

PA (AREN-) ARENA PHARM INC.

XX

PI Liaw CW, Behan DP, Chalmers DT;

XX

WPI; 2003-742861/70.

XX

N-PSDB; ADC22796.

XX

PT Creating a constitutively active version of an endogenous human G protein
 PT coupled receptor (GPCR) comprises substituting a specific amino acid in
 PT the transmembrane-6 region with a different amino acid, and testing for
 PT constitutive activity.

PS Example 2; SEQ ID NO 278; 221pp; English.

XX The invention relates to a method for treating a non-endogenous,
 CC constitutively active version of an endogenous human G protein-coupled
 CC receptor (GPCR) that has a transmembrane-6 (TM6) region and an
 CC intracellular-3 (IC3) region, by substituting a specific amino acid in
 CC the TM6 region with a different amino acid, and testing for constitutive
 CC activity. The method is useful for creating a constitutively active
 CC version of an endogenous human GPCR that comprises a transmembrane 6
 CC region and an intracellular loop 3 region. The altered human GPCR
 CC polypeptides are useful for screening test compounds for identification
 CC of inverse agonists or partial agonists of GPCR polypeptides, which may
 CC have therapeutic uses. The altered GPCRs may also be used in vivo or in
 CC vitro in biological research. A nucleic acid encoding the altered GPCR
 CC may be used to create a transgenic animal expressing the altered GPCR.
 CC The method allows screening for compounds that modulate the activity of a
 CC human G protein-coupled receptor without the need for provision of a
 CC ligand for the receptor. This is particularly useful in allowing
 CC screening of compounds against orphan receptors for which no ligand is
 CC currently known. This sequence represents a human GPCR polypeptide of the
 CC invention.

XX Sequence 907 AA;

Query Match 0.9%; Score 9; DB 7; Length 907;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
 Db 174 LTEIPVQAF 182
 |||||

RESULT 36

ADC22783

ID ADC22783 standard; protein; 907 AA.

XX AC ADC22783;

XX DT 18-DEC-2003 (first entry)

XX DE Human G protein-coupled receptor (GPCR) polypeptide #39.

XX KW Human; G protein-coupled receptor; GPCR; transmembrane-6 region; TM6;
 KW intracellular-3 region; IC3; receptor.

XX OS Homo sapiens.

XX PN US6555339-B1.

XX PD 29-APR-2003.

XX PF 13-OCT-1998; 98US-00170496.

XX PR 14-APR-1997; 97US-00839449.

XX PR 14-APR-1998; 98US-00060188.

XX PR 26-JUN-1998; 98US-0090783P.

XX PR 07-AUG-1998; 98US-0095677P.

XX PA (AREN-) ARENA PHARM INC.

XX PI Liaw CW, Behan DF, Chalmers DT;

XX DR WPI; 2003-742861/70.

XX DR N-PSDB; ADC22782.

XX PT Creating a constitutively active version of an endogenous human G protein
 PT coupled receptor (GPCR) comprises substituting a specific amino acid in

PT the transmembrane-6 region with a different amino acid, and testing for
 PT constitutive activity.

PS Example 1; SEQ ID NO 264; 221pp; English.

XX The invention relates to a method for treating a non-endogenous,
 CC constitutively active version of an endogenous human G protein-coupled
 CC receptor (GPCR) that has a transmembrane-6 (TM6) region and an
 CC intracellular-3 (IC3) region, by substituting a specific amino acid in
 CC the TM6 region with a different amino acid, and testing for constitutive
 CC activity. The method is useful for creating a constitutively active
 CC version of an endogenous human GPCR that comprises a transmembrane 6
 CC region and an intracellular loop 3 region. The altered human GPCR
 CC polypeptides are useful for screening test compounds for identification
 CC of inverse agonists or partial agonists of GPCR polypeptides, which may
 CC have therapeutic uses. The altered GPCRs may also be used in vivo or in
 CC vitro in biological research. A nucleic acid encoding the altered GPCR
 CC may be used to create a transgenic animal expressing the altered GPCR.
 CC The method allows screening for compounds that modulate the activity of a
 CC human G protein-coupled receptor without the need for provision of a
 CC ligand for the receptor. This is particularly useful in allowing
 CC screening of compounds against orphan receptors for which no ligand is
 CC currently known. This sequence represents a human GPCR polypeptide of the
 CC invention.

XX Sequence 907 AA;

Query Match 0.9%; Score 9; DB 7; Length 907;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
 Db 174 LTEIPVQAF 182
 |||||

RESULT 37

ADE59150

ID ADE59150 standard; protein; 907 AA.

XX AC ADE59150;

XX DT 29-JAN-2004 (first entry)

XX DE Human Protein O75473, SEQ ID NO 5041.

XX KW Human; pain; neuronal tissue; gene therapy;

XX KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.

XX OS Homo sapiens.

XX PN WO2003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GEHO) GEN HOSPITAL CORP.

XX PA (FARB) BAYER AG.

XX PI Woolf C, D'urso D, Befort K, Costigan M;

XX DR WPI; 2003-268312/26.

XX DR GENBANK; O75473.

XX PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat

CC or human polynucleotides or a polynucleotide which represents a fragment,

CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell

CC comprising the vector, a method for identifying a nucleotide sequence

CC which is differentially regulated in an animal subjected to pain and a

CC kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence

CC that is differentially expressed in neuronal tissue of a first animal

CC subjected to pain, a method for identifying a compound which regulates

CC the expression of a polynucleotide sequence which is differentially

CC expressed in an animal subjected to pain, a method for identifying a

CC compound that regulates the activity of one or more of the

CC polynucleotides, a method for producing a pharmaceutical composition, a

CC method for identifying a compound or small molecule that regulates the

CC activity in an animal of one or more of the polypeptides given in the

CC specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more

CC polypeptides or their antibodies. The polynucleotide or the compound that

CC modulates its activity is useful for preparing a medicament for treating

CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene

CC therapy). The sequence presented is a human protein (shown in Table 2 of

CC the specification) which is differentially expressed during pain. Note:

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic form directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 907 AA;

QY 592 LTEIPVQAF 600

DB 174 LTEIPVQAF 182

RESULT 38

AD59153

ID AD59153 standard; protein; 907 AA.

XX AC AD59153;

XX 29-JAN-2004 (first entry)

XX Human Protein O75473, SEQ ID NO 5044.

XX Human; pain; neuronal tissue; gene therapy;

XX spinal segmental nerve injury; chronic constriction injury; CCI;

XX spared nerve injury; SNI; Chung.

XX Homo sapiens.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

XX 01-NOV-2001; 2001US-0346382P.

XX 26-NOV-2001; 2001US-0333347P.

XX (GEO) GEN HOSPITAL CORP.

XX (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

GENBANK; O75473.

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 907 AA;

Query Match 0.9%; Score 9; DB 7; Length 907;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600

DB 174 LTEIPVQAF 182

RESULT 39

ADG42629

ID ADG42629 standard; protein; 907 AA.

XX AC ADG42629;

XX 26-FEB-2004 (first entry)

XX Human G protein-coupled receptor LGR5.

XX cytostatic; gene therapy; NOX-agonist; NOX-antagonist; pharmaceutical;

XX NOX-associated disorder; cancer; human; G protein coupled receptor LGR5.

XX Homo sapiens.

XX US2003204052-A1.

XX 30-OCT-2003.

XX 04-OCT-2001; 2001US-00970944.

XX 04-OCT-2000; 2000US-0237862P.

XX (HERR) HERRMANN J L.

XX (RAST) RASTELLI L.

XX (SHIM) SHIMKETS R A.

PI Herrmann JL, Rastelli L, Shinkets RA;
XX WPI; 2003-900673/82.
XX New NOVX gene or NOVX-specific antibody, useful for preparing a
PT composition for treating or preventing a NOVX-associated disorder, e.g.,
PT cancer.
XX Disclosure; SEQ ID NO 27; 118pp; English.
XX The invention describes a new isolated polypeptide comprising: a
CC polypeptide or its mature form comprising a sequence not given in the
CC specification; or a variant of (A), where one or more amino acid residues
CC in the variant differs in no more than 15% from the amino acid sequence
CC of the mature form. The pharmaceutical composition may be administered
CC via oral, transdermal, rectal or parenteral route. The polypeptide,
CC nucleic acid or antibody is useful for preparing a composition for
CC treating or preventing a NOVX-associated disorder, e.g., cancer. This is
CC the amino acid sequence of a transmembrane receptor homologue used in a
CC comparison with the novel human proteins of the invention.
XX Sequence 907 AA;
SQ

Query Match 0.9%; Score 9; DB 7; Length 907;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 592 LTEIPVQAF 600
DB 174 LTEIPVQAF 182
|||||

RESULT 40
ADG42628
ID ADG42628 standard; protein; 907 AA.
XX AC ADG42628;
XX 26-FEB-2004 (first entry)
XX Human G protein-coupled receptor 49.
XX cytostatic; gene therapy; NOVX-agonist; NOVX-antagonist; pharmaceutical;
XX NOVX-associated disorder; cancer; human; G protein coupled receptor 49.
XX Homo sapiens.
XX US2003204052-A1.
XX 30-OCT-2003.
XX 04-OCT-2001; 2001US-00970944.
XX 04-OCT-2000; 2000US-0237862P.
XX (HERR/) HERRMANN J L.
XX (RAST/) RASTELLI L.
XX (SHIM/) SHINKETS R A.
PI Herrmann JL, Rastelli L, Shinkets RA;
XX WPI; 2003-900673/82.
XX New NOVX gene or NOVX-specific antibody, useful for preparing a
PT composition for treating or preventing a NOVX-associated disorder, e.g.,
PT cancer.
XX Disclosure; SEQ ID NO 26; 118pp; English.
XX The invention describes a new isolated polypeptide comprising: a
CC polypeptide or its mature form comprising a sequence not given in the
CC specification; or a variant of (A), where one or more amino acid residues
CC in the variant differs in no more than 15% from the amino acid sequence

CC of the mature form. The pharmaceutical composition may be administered
CC via oral, transdermal, rectal or parenteral route. The polypeptide,
CC nucleic acid or antibody is useful for preparing a composition for
CC treating or preventing a NOVX-associated disorder, e.g., cancer. This is
CC the amino acid sequence of a transmembrane receptor homologue used in a
CC comparison with the novel human proteins of the invention.
XX Sequence 907 AA;
SQ

Query Match 0.9%; Score 9; DB 7; Length 907;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 592 LTEIPVQAF 600
DB 174 LTEIPVQAF 182
|||||

RESULT 41
ADH14270
ID ADH14270 standard; protein; 907 AA.
XX AC ADH14270;
XX 11-MAR-2004 (first entry)
XX Mutated human HG38.
XX human; non-endogenous; G protein-coupled receptor; GPCR; receptor;
XX mutant; mutein.
XX Synthetic.
XX Homo sapiens.
XX US2003105292-A1.
XX 05-JUN-2003.
XX 20-SEP-2002; 2002US-00251385.
XX 26-JUN-1998; 98US-0090783P.
XX 07-AUG-1998; 98US-0095677P.
XX 13-OCT-1998; 98US-00170496.
XX (LIAW/) LIAW C W.
XX (BEHA/) BEHAN D P.
XX (CHAL/) CHALMERS D T.
PI Liaw CW, Behan DP, Chalmers DT;
XX WPI; 2003-801247/75.
XX N-PSDB; ADH14269.
XX New constitutively active, non-endogenous version of an endogenous human
XX G protein-coupled receptor for the identification of therapeutic
XX compounds, such as agonists.
XX Example 2; SEQ ID NO 278; 227pp; English.
XX The invention relates to a constitutively active, non-endogenous version
XX of an endogenous human G protein-coupled receptor (GPCR). The GPCR is
XX used for screening therapeutic compounds as inverse agonists, agonists or
XX partial agonists. The GPCR can be also be used to elucidate and
XX understand the roles of GPCRs in normal and diseased humans. The GPCR
XX need not be purified and isolated to be used to screen for therapeutic
XX compounds. The utility of the GPCR as a research tool is enhanced because
XX the role of a particular receptor can be understood before the endogenous
XX ligand is identified. The present sequence is used in the exemplification
XX of the present invention.
SQ Sequence 907 AA;
Query Match 0.9%; Score 9; DB 7; Length 907;

Best Local Similarity 100.0%; Pred. No. 45;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 42
ADH14256
ID ADH14256 standard; protein; 907 AA.
AC ADH14256;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human HG38.
XX
KW human; non-endogenous; G protein-coupled receptor; GPCR; receptor.
XX
OS Homo sapiens.
XX
FN US2003105292-A1.
XX
PD 05-JUN-2003.
XX
PF 20-SEP-2002; 2002US-00251385.
XX
PR 26-JUN-1998; 98US-0090783P.
PR 07-AUG-1998; 98US-0095677P.
PR 13-OCT-1998; 98US-00170496.
XX
PA (LIAM/) LIAM C W.
PA (BEHA/) BEHA D P.
PA (CHAL/) CHALMERS D T.
XX
PI Liaw CW, Behan DP, Chalmers DT;
XX
XX WPI; 2003-801247/75.
DR N-PSDB; ADH14255.
XX
XX New constitutively active, non-endogenous version of an endogenous human
PT G protein-coupled receptor for the identification of therapeutic
PT compounds, such as agonists.
XX
PS Example 1; SEQ ID NO 264; 227pp; English.
XX
XX The invention relates to a constitutively active, non-endogenous version
CC of an endogenous human G protein-coupled receptor (GPCR). The GPCR is
CC used for screening therapeutic compounds as inverse agonists, agonists or
CC partial agonists. The GPCR can be also be used to elucidate and
CC understand the roles of GPCRs in normal and diseased humans. The GPCR
CC need not be purified and isolated to be used to screen for therapeutic
CC compounds. The utility of the GPCR as a research tool is enhanced because
CC the role of a particular receptor can be understood before the endogenous
CC ligand is identified. The present sequence is used in the exemplification
XX of the present invention.
XX
XX Sequence 907 AA;
XX

Query Match 0.9%; Score 9; DB 7; Length 907;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 43
ADN40013
ID ADN40013 standard; protein; 907 AA.
XX

AC ADN40013;
XX
DT 17-JUN-2004 (first entry)
XX
DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:C383.
XX
DE Human; differential expression; cancer; angiogenic disorder;
KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;
KW inflammatory disease; autoimmune disease;
KW retinal neovascularisation syndrome; scarring; uterine fibroid;
KW detection; diagnosis; prognosis; drug screening; drug targeting;
KW wound healing; contraception; cytostatic; cardiac; immunomodulatory;
KW vulnerary; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
FN WO2003042661-A2.
XX
PD 22-MAY-2003.
XX
PF 13-NOV-2002; 2002WO-US036810.
XX
XX 13-NOV-2001; 2001US-0350666P.
PR 21-NOV-2001; 2001US-0332464P.
PR 29-NOV-2001; 2001US-0334393P.
PR 03-DEC-2001; 2001US-0335394P.
PR 14-DEC-2001; 2001US-0340376P.
PR 08-JAN-2002; 2002US-0347211P.
PR 10-JAN-2002; 2002US-0347349P.
PR 08-FEB-2002; 2002US-0355250P.
PR 13-FEB-2002; 2002US-0356714P.
PR 20-FEB-2002; 2002US-0359077P.
PR 29-MAR-2002; 2002US-0368809P.
PR 04-APR-2002; 2002US-0370110P.
PR 12-APR-2002; 2002US-0372246P.
PR 05-JUN-2002; 2002US-0386614P.
PR 16-JUL-2002; 2002US-0396839P.
PR 22-JUL-2002; 2002US-0397775P.
PR 22-JUL-2002; 2002US-0397845P.
PR 09-SEP-2002; 2002US-0409450P.
XX
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;
PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;
XX
XX WPI; 2003-468649/44.
DR N-PSDB; ADN39796.
XX
XX Determining the presence or absence of a pathological cell in a patient,
PT useful for diagnosing, prognosing or treating cancer, comprises detecting
PT a nucleic acid in a biological sample.
XX
PS Claim 12; SEQ ID NO C383; 1385pp; English.
XX
XX The invention relates to nucleic acids and proteins (ADN38683-ADN40064)
CC whose expression is upregulated or downregulated in specific cancers or
CC other diseases such as angiogenic or fibrotic disorders, and to methods
CC of determining the presence or absence of a pathological cell in a
CC patient by detecting a nucleic acid at least 80% identical to those of
CC the invention or by detecting a polypeptide of the invention. The
CC invention also relates to expression vectors and host cells comprising a
CC nucleic acid of the invention; antibodies which specifically bind a
CC polypeptide of the invention; use of such antibodies for drug targeting;
CC and methods of screening for modulators of activity or expression of the
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,
CC antibodies and methods are useful for diagnosing, prognosing and treating
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal
CC neovascularisation syndromes, scarring and uterine fibroids. They may
CC also be useful in wound healing and in contraception. The present
CC sequence represents a polypeptide of the invention.
XX

SQ Sequence 907 AA;
 Query Match 0.9%; Score 9; DB 7; Length 907;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
 DB 174 LTEIPVQAF 182
 |||||

RESULT 44
 ADN39531
 ID ADN39531 standard; protein; 907 AA.
 XX AC ADN39531;
 XX DT 17-JUN-2004 (first entry)
 XX DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:A131.
 XX KW Human; differential expression; cancer; angiogenic disorder;
 KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;
 KW inflammatory disease; autoimmune disease;
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;
 KW detection; diagnosis; prognosis; drug screening; drug targeting;
 KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;
 KW vulnery; gene therapy; vaccine.
 XX OS Homo sapiens.
 XX PN WO2003042661-A2.
 XX PD 22-MAY-2003.
 XX PP 13-NOV-2002; 2002WO-US036810.
 XX PR 13-NOV-2001; 2001US-0350666P.
 PR 21-NOV-2001; 2001US-0332464P.
 PR 29-NOV-2001; 2001US-0334393P.
 PR 03-DEC-2001; 2001US-0335394P.
 PR 14-DEC-2001; 2001US-0340376P.
 PR 08-JAN-2002; 2002US-0347211P.
 PR 10-JAN-2002; 2002US-0347349P.
 PR 08-FEB-2002; 2002US-0355250P.
 PR 13-FEB-2002; 2002US-0356714P.
 PR 20-FEB-2002; 2002US-0359077P.
 PR 29-MAR-2002; 2002US-0368099P.
 PR 04-APR-2002; 2002US-0370110P.
 PR 12-APR-2002; 2002US-0372246P.
 PR 05-JUN-2002; 2002US-0386614P.
 PR 16-JUL-2002; 2002US-0396839P.
 PR 22-JUL-2002; 2002US-0397775P.
 PR 22-JUL-2002; 2002US-0397845P.
 PR 09-SEP-2002; 2002US-0409450P.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 XX AFAR D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Heverzi PA;
 PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;
 XX WPI; 2003-468649/44.
 DR N-PSDB; ADN39530.
 XX
 PT Determining the presence or absence of a pathological cell in a patient,
 PT useful for diagnosing, prognosing or treating cancer, comprises detecting
 PT a nucleic acid in a biological sample.
 XX
 PS Claim 12; SEQ ID NO A131; 1385pp; English.
 XX
 CC The invention relates to nucleic acids and proteins (ADN38683-ADN40064)
 CC whose expression is upregulated or downregulated in specific cancers or
 CC other diseases such as angiogenic or fibrotic disorders, and to methods

CC of determining the presence or absence of a pathological cell in a
 CC patient by detecting a nucleic acid at least 80% identical to those of
 CC the invention or by detecting a polypeptide of the invention. The
 CC invention also relates to expression vectors and host cells comprising a
 CC nucleic acid of the invention; antibodies which specifically bind a
 CC polypeptide of the invention; use of such antibodies for drug targeting;
 CC and methods of screening for modulators of activity or expression of the
 CC polypeptides and nucleic acids. The nucleic acids, polypeptides,
 CC antibodies and methods are useful for diagnosing, prognosing and treating
 CC cancer and other conditions such as psoriasis, ischaemia, heart disease,
 CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal
 CC neovascularisation syndromes, scarring and uterine fibroids. They may
 CC also be useful in wound healing and in contraception. The present
 CC sequence represents a polypeptide of the invention.
 XX
 SQ Sequence 907 AA;
 Query Match 0.9%; Score 9; DB 7; Length 907;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
 DB 174 LTEIPVQAF 182
 |||||

RESULT 45
 ADN39628
 ID ADN39628 standard; protein; 907 AA.
 XX AC ADN39628;
 XX DT 17-JUN-2004 (first entry)
 XX DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:A228.
 XX KW Human; differential expression; cancer; angiogenic disorder;
 KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;
 KW inflammatory disease; autoimmune disease;
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;
 KW detection; diagnosis; prognosis; drug screening; drug targeting;
 KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;
 KW vulnery; gene therapy; vaccine.
 XX OS Homo sapiens.
 XX PN WO2003042661-A2.
 XX PD 22-MAY-2003.
 XX PP 13-NOV-2002; 2002WO-US036810.
 XX PR 13-NOV-2001; 2001US-0350666P.
 PR 21-NOV-2001; 2001US-0332464P.
 PR 29-NOV-2001; 2001US-0334393P.
 PR 03-DEC-2001; 2001US-0335394P.
 PR 14-DEC-2001; 2001US-0340376P.
 PR 08-JAN-2002; 2002US-0347211P.
 PR 10-JAN-2002; 2002US-0347349P.
 PR 08-FEB-2002; 2002US-0355250P.
 PR 13-FEB-2002; 2002US-0356714P.
 PR 20-FEB-2002; 2002US-0359077P.
 PR 29-MAR-2002; 2002US-0368099P.
 PR 04-APR-2002; 2002US-0370110P.
 PR 12-APR-2002; 2002US-0372246P.
 PR 05-JUN-2002; 2002US-0386614P.
 PR 16-JUL-2002; 2002US-0396839P.
 PR 22-JUL-2002; 2002US-0397775P.
 PR 22-JUL-2002; 2002US-0397845P.
 PR 09-SEP-2002; 2002US-0409450P.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 XX AFAR D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Heverzi PA;
 PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;
 XX WPI; 2003-468649/44.
 DR N-PSDB; ADN39530.
 XX
 PT Determining the presence or absence of a pathological cell in a patient,
 PT useful for diagnosing, prognosing or treating cancer, comprises detecting
 PT a nucleic acid in a biological sample.
 XX
 PS Claim 12; SEQ ID NO A131; 1385pp; English.
 XX
 CC The invention relates to nucleic acids and proteins (ADN38683-ADN40064)
 CC whose expression is upregulated or downregulated in specific cancers or
 CC other diseases such as angiogenic or fibrotic disorders, and to methods

PI Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;
 PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;
 XX WPI; 2003-468649/44.
 DR N-PSDB; ADN39627.
 XX
 PT Determining the presence or absence of a pathological cell in a patient,
 PT useful for diagnosing, prognosing or treating cancer, comprises detecting
 PT a nucleic acid in a biological sample.
 XX
 PS Claim 12; SEQ ID NO A228; 1385pp; English.
 XX
 CC The invention relates to nucleic acids and proteins (ADN38683-ADN40064)
 CC whose expression is upregulated or downregulated in specific cancers or
 CC other diseases such as angiogenic or fibrotic disorders, and to methods
 CC of determining the presence or absence of a pathological cell in a
 CC patient by detecting a nucleic acid at least 80% identical to those of
 CC the invention or by detecting a polypeptide of the invention. The
 CC invention also relates to expression vectors and host cells comprising a
 CC nucleic acid of the invention; antibodies which specifically bind a
 CC polypeptide of the invention; use of such antibodies for drug targeting;
 CC and methods of screening for modulators of activity or expression of the
 CC polypeptides and nucleic acids. The nucleic acids, polypeptides,
 CC antibodies and methods are useful for diagnosing, prognosing and treating
 CC cancer and other conditions such as psoriasis, ischaemia, heart disease,
 CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal
 CC neurovascularisation syndromes, scarring and uterine fibroids. They may
 CC also be useful in wound healing and in contraception. The present
 CC sequence represents a polypeptide of the invention.
 XX
 SQ Sequence 907 AA;
 Query Match 0.9%; Score 9; DB 7; Length 907;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 592 LTEIPVQAF 600
 Db 174 LTEIPVQAF 182
 |||||
 RESULT 46
 ID ADO29408
 XX ADO29408 standard; protein; 907 AA.
 XX ADO29408;
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE Human GPCR GPR49, SEQ ID NO:510.
 XX
 KW G protein-coupled receptor; GPCR; drug screening; diagnosis;
 KW transgenic mouse; neurological disorder; adrenal gland disorder;
 KW colon disorder; intestinal disorder; cardiovascular disorder;
 KW muscular disorder; blood disorder; immune disorder; bone disorder;
 KW joint disorder; metabolic disorder; nutritive disorder; cancer;
 KW kidney disorder; liver disorder; lung disorder; breast disorder;
 KW ovary disorder; uterus disorder; prostate disorder; testis disorder;
 KW skin disorder; stomach disorder; pancreas disorder; spleen disorder;
 KW thymus disorder; thyroid disorder; antiparkinsonian; antimanic;
 KW cystostatic; antiinflammatory; vasotropic; antiangiinal; antiarrhythmic;
 KW CNS; central nervous system; respiratory; antidiarrhoeic; antidiabetic;
 KW virucide; hepatotropic; antibacterial; antianemic; antieborrheic;
 KW dermatological; antitumor; antithyroid; antiallergic; anorectic;
 KW immunosuppressive; nephrotropic; gene therapy; GPCR modulator; human;
 KW receptor.
 XX
 OS Homo sapiens.
 XX
 PN WO2004040000-A2.
 XX
 PD 13-MAY-2004.
 XX

PP 09-SEP-2003; 2003WO-US028226.
 XX
 PR 09-SEP-2002; 2002US-0409303P.
 PR 09-APR-2003; 2003US-0461329P.
 XX
 PA (PRIM-) PRIMAL INC.
 XX
 PI Gaitanaris GA, Bergmann JE, Gragerov A, Hohmann J, Li F;
 PI Madisen L, McIlwain KL, Pavlova MN, Vassiliadis D, Zeng H;
 XX
 DR WPI; 2004-390329/36.
 DR N-PSDB; ADO29923.
 XX
 PT Novel mammalian G protein coupled receptors, useful for identifying
 PT compounds that modulates diagnosing and treating disease condition
 PT associated with GPCR dysfunction e.g. autoimmune diseases, angina
 PT pectoris, Parkinson's disease.
 XX
 PS Claim 151; SEQ ID NO 510; 542pp; English.
 XX
 CC The invention relates to human and mouse G protein-coupled receptors
 CC (GPCRs) and nucleic acids encoding them. The invention also relates to
 CC sequences at least 90% identical to the GPCR proteins and nucleic acids
 CC of the invention; methods of treating, preventing or diagnosing diseases
 CC associated with GPCRs of the invention; methods of screening for
 CC compounds useful in the treatment of GPCR-related diseases; a transgenic
 CC mouse comprising a GPCR gene of the invention; a mouse comprising a
 CC mutation in a GPCR transgene or in an endogenous GPCR gene; cells derived
 CC from the transgenic mice; kits comprising several mice, each of which has
 CC a mutation in a different GPCR gene of the invention; and kits comprising
 CC probes which hybridise to GPCR polynucleotides of the invention. The
 CC invention further discloses variants of the GPCR polypeptides and vectors
 CC comprising a GPCR nucleic acid. The GPCR nucleic acids and proteins may
 CC be used in the diagnosis, treatment or prevention of a wide variety of
 CC diseases including neurological disorders (e.g., Alzheimer's disease,
 CC depression, diabetic neuropathy, Parkinson's disease or schizophrenia);
 CC disorders of the adrenal gland; disorders of the colon or intestine
 CC (e.g., Crohn's disease, diarrhoea, food poisoning or irritable bowel
 CC syndrome); cardiovascular disorders (e.g., angina, cardiac arrhythmia or
 CC myocardial infarction); muscular disorders; blood disorders (e.g.,
 CC anaemia or leukaemia); immune disorders (e.g., autoimmune disorders or
 CC AIDS); bone and joint disorders (e.g., osteoarthritis, rheumatoid
 CC arthritis, gout or osteoporosis); metabolic or nutritive disorders (e.g.,
 CC obesity, enzyme deficiency-related diseases or vitamin deficiency-related
 CC diseases); and disorders of the kidney, liver, lung, breast, ovary,
 CC uterus, prostate, testis, skin, stomach, pancreas, spleen, thymus and
 CC thyroid (e.g., cancers). The present sequence represents a GPCR of the
 CC invention. Note: The full sequence data for this patent did not form part
 CC of the printed specification; those sequences not shown were obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 907 AA;
 Query Match 0.9%; Score 9; DB 8; Length 907;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 592 LTEIPVQAF 600
 Db 174 LTEIPVQAF 182
 |||||
 RESULT 47
 ID ADO80369
 XX ADO80369 standard; protein; 907 AA.
 XX ADO80369;
 XX
 DT 21-OCT-2004 (first entry)
 XX
 DE G protein-coupled receptor 49 protein.
 XX

PI Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
 XX WPI; 2003-697654/66.
 DR N-PSDB; ADF70582.
 XX Transformation of cells with a fusion protein of an orphan receptor
 PT protein with a fluorescent protein useful for identification of ligands
 PT to the orphan receptor.
 XX Disclosure; SEQ ID NO 103; 594pp; Japanese.
 XX This invention relates to a novel method of identifying ligands to an
 CC orphan receptor protein which comprises transforming cells with DNA
 CC encoding a fusion protein of the orphan receptor with a fluorescent
 CC protein, so that the fusion protein is expressed in the cells (or cell
 CC membranes isolated from them) and contacting the cells with the potential
 CC ligand to be tested. A suitable fluorescent protein for incorporation in
 CC the fusion protein is green fluorescent protein (GFP), for example GFP-1,
 CC wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
 CC identification of ligands binding to an orphan receptor protein.
 XX

SQ Sequence 1145 AA;
 Query Match 0.9%; Score 9; DB 7; Length 1145;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 592 LTFIPVQAF 600
 Db 174 LTFIPVQAF 182
 |||||
 |||||

RESULT 50
 ABG03518
 ID ABG03518 standard; protein; 147 AA.
 XX AC ABG03518;
 XX DT 13-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #3509.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens!
 OS WO200175067-A2.
 PN 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US008631.
 XX PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS67705.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID NO 33877; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: the sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 147 AA;
 Query Match 0.8%; Score 8; DB 4; Length 147;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 983 SFVYGILG 990
 Db 123 SFVYGILG 130
 |||||
 |||||

RESULT 51
 ADG88441
 ID ADG88441 standard; protein; 196 AA.
 XX AC ADG88441;
 XX DT 11-MAR-2004 (first entry)
 XX DE Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANTND4.
 XX Plant; genetic engineering; cell proliferation; aintegumenta; ANT;
 KW transgenic; transgenic plant; organ mass alteration; fertility;
 KW asexual reproduction; mutant; muten.
 XX OS Synthetic.
 OS Arabidopsis thaliana.
 PN US2003159180-A1.
 XX PD 21-AUG-2003.
 XX PF 28-JAN-2002; 2002US-00059911.
 XX PR 28-JAN-2002; 2002US-00059911.
 XX PA (REGC) UNIV CALIFORNIA OFFICE TECHNOLOGY.
 XX PI Fischer RL, Mizukami Y;
 XX WPI; 2003-787370/74.
 DR N-PSDB; ADG88458.
 XX Increasing cell proliferation in a plant by introducing into the plant an
 PT expression cassette containing a plant promoter operably linked to a
 PT nucleic acid encoding a modified ANT polypeptide and selecting plants
 PT with increased mass.
 XX Disclosure; SEQ ID NO 12; 51pp; English.
 XX The present invention relates to plant genetic engineering. The invention
 CC particularly relates to a method of increasing cell proliferation in a
 CC plant. The method involves introducing into the plant an expression
 CC cassette containing a plant promoter operably linked to a nucleic acid

CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2
 CC domain and selecting plants the with increased size or mass. The
 CC invention is useful to generate transgenic plants. The method is useful
 CC for increasing cell proliferation in a plant for altering organ mass,
 CC controlling fertility or enhancing asexual reproduction. The present
 CC sequence is Arabidopsis thaliana ANT mutant protein. This sequence is
 CC used in the exemplification of the invention.

XX SQ Sequence 196 AA;

Query Match 0.8%; Score 8; DB 7; Length 196;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LKTVALTL 12
 |||||
 DB 177 LKTVALTL 184

RESULT 52

ADM26567
 ID ADM26567 standard; protein; 240 AA.

XX AC ADM26567;

XX DT 20-MAY-2004 (first entry)

XX DE Hyperthermophile Methanopyrus kandleri protein #1173.

XX KW hyperthermophile; protein stability enhancement;

XX KW protein activity enhancement.

XX OS Methanopyrus kandleri.

XX PN WO2003076575-A2.

XX PD 18-SEP-2003.

XX PF 04-MAR-2003; 2003WO-US006664.

XX PR 04-MAR-2002; 2002US-0361742P.

XX PR 14-MAY-2002; 2002US-0380423P.

XX PR 16-SEP-2002; 2002US-0410974P.

XX (FIDE-) FIDELITY SYSTEMS INC.

XX PA (MALY/) MALYKH A.

XX PI Slesarev AI, Pavlov A, Pavlova N, Kozyavkin S;

XX DR WPI; 2003-748383/70.

XX DR N-PSDB; ADM27081.

XX PT New isolated nucleic acids encoding any of about 1700 Methanopyrus
 PT kandleri proteins, and the encoded proteins, useful as a medicaments or
 PT as diagnostic agents.

XX PS Claim 31; SEQ ID NO 1173; 1023pp; English.

XX CC The invention comprises the amino acid sequence of proteins from the
 CC hyperthermophile Methanopyrus kandleri, the invention also comprises the
 CC complete genome from Methanopyrus kandleri. The Methanopyrus kandleri
 CC proteins of the invention are useful for enhancing the stability and/or
 CC activity of other proteins. The Methanopyrus kandleri genome is useful in
 CC a variety of diagnostic and analytical methods. The present amino acid
 CC sequence represents a Methanopyrus kandleri protein of the invention.

XX SQ Sequence 240 AA;

Query Match 0.8%; Score 8; DB 7; Length 240;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 794 RLEKELDL 801

DB 44 RLEKELDL 51
 |||||

RESULT 53

ADG88453
 ID ADG88453 standard; protein; 264 AA.

XX AC ADG88453;

XX DT 11-MAR-2004 (first entry)

XX DE Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANT-AP2sw2.

XX KW Plant; genetic engineering; cell proliferation; aintegumenta; ANT;

XX KW transgenic; transgenic plant; organ mass alteration; fertility;

XX KW asexual reproduction; mutant; mutein.

XX OS Synthetic.

XX OS Arabidopsis thaliana.

XX PN US2003159180-A1.

XX PD 21-AUG-2003.

XX PF 28-JAN-2002; 2002US-00059911.

XX PR 28-JAN-2002; 2002US-00059911.

XX PA (REGC) UNIV CALIFORNIA OFFICE TECHNOLOGY.

XX PI Fischer RL, Mizukami Y;

XX DR WPI; 2003-787370/74.

XX DR N-PSDB; ADG8470.

XX PT Increasing cell proliferation in a plant by introducing into the plant an
 PT expression cassette containing a plant promoter operably linked to a
 PT nucleic acid encoding a modified ANT polypeptide and selecting plants
 PT with increased mass.

XX PS Disclosure; SEQ ID NO 24; 51pp; English.

XX CC The present invention relates to plant genetic engineering. The invention
 CC particularly relates to a method of increasing cell proliferation in a
 CC plant. The method involves introducing into the plant an expression
 CC cassette containing a plant promoter operably linked to a nucleic acid
 CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2
 CC domain and selecting plants the with increased size or mass. The
 CC invention is useful to generate transgenic plants. The method is useful
 CC for increasing cell proliferation in a plant for altering organ mass,
 CC controlling fertility or enhancing asexual reproduction. The present
 CC sequence is Arabidopsis thaliana ANT mutant protein. This sequence is
 CC used in the exemplification of the invention.

XX SQ Sequence 264 AA;

Query Match 0.8%; Score 8; DB 7; Length 264;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LKTVALTL 12

DB 245 LKTVALTL 252

RESULT 54

ABB05713
 ID ABB05713 standard; protein; 268 AA.

XX AC ABB05713;

XX DT 30-APR-2002 (first entry)

XX DE Human cell structure and motility protein clone tes3_16b5.
 XX KW Human; foetal brain; foetal kidney; melanoma; testis; amygdala;
 XX KW gene therapy.
 XX OS Homo sapiens.
 XX PN WO200198454-A2.
 XX PD 27-DEC-2001.
 XX PF 25-APR-2001; 2001WO-IB002050.
 XX PR 25-APR-2000; 2000US-0199380P.
 XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
 XX PI Wiemann S;
 XX XN WPI; 2002-055860/07.
 XX DR N-PSDB; ABA93750.
 XX PT Human cDNA sequences and clones derived from human fetal brain, fetal
 PT kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic
 PT screening and therapy.
 XX PS Claim 1; Page 315; 61lpp; English.
 XX CC The present invention describes assemblages and computer readable media
 CC comprising novel human cDNA sequences and clones derived from human
 CC foetal brain, foetal kidney, melanoma, testis and amygdala cDNA
 CC libraries. ABA93702 to ABA93766 represent human cDNA sequences from the
 CC present invention which encode the proteins given in ABB05662 to
 CC ABB05729. The human cDNA sequences and clones can be used in gene
 CC therapy. The clones may be used in a variety of applications, for example
 CC they may be used in profiling assays, for providing large arrays of human
 CC genetic material for implementing large-scale screening strategies and
 CC for treating diseases via gene therapy procedures
 XX SQ Sequence 268 AA;
 Query Match 0.8%; Score 8; DB 5; Length 268;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 793 KELEKELD 800
 DB 83 KELEKELD 90
 RESULT 55
 ADG88454
 ID ADG88454 standard; protein; 275 AA.
 XX AC ADG88454;
 XX DT 11-MAR-2004 (first entry)
 XX DE Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANTdf1.
 XX KW Plant; genetic engineering; cell proliferation; aintegumenta; ANT;
 KW transgenic; transgenic plant; organ mass alteration; fertility;
 KW asexual reproduction; mutant; muten.
 XX OS Synthetic.
 OS Arabidopsis thaliana.
 XX Key Location/Qualifiers
 FH Misc-difference 1..2
 FT /note= "Encoded by ATGACTTCT"
 FT Misc-difference 134
 FT /note= "Encoded by GGA"

XX PN US2003159180-A1.
 XX PD 21-AUG-2003.
 XX PF 28-JAN-2002; 2002US-00059911.
 XX PR 28-JAN-2002; 2002US-00059911.
 XX PA (REGC) UNIV CALIFORNIA OFFICE TECHNOLOGY.
 XX PI Fischer RL, Mizukami Y;
 XX XN WPI; 2003-787370/74.
 XX DR N-PSDB; ADG88471.
 XX PT Increasing cell proliferation in a plant by introducing into the plant an
 PT expression cassette containing a plant promoter operably linked to a
 PT nucleic acid encoding a modified ANT polypeptide and selecting plants
 PT with increased mass.
 XX PS Claim 4; SEQ ID NO 25; 5lpp; English.
 XX CC The present invention relates to plant genetic engineering. The invention
 CC particularly relates to a method of increasing cell proliferation in a
 CC plant. The method involves introducing into the plant an expression
 CC cassette containing a plant promoter operably linked to a nucleic acid
 CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2
 CC domain and selecting plants the with increased size or mass. The
 CC invention is useful to generate transgenic plants. The method is useful
 CC for increasing cell proliferation in a plant for altering organ mass,
 CC controlling fertility or enhancing asexual reproduction. The present
 CC sequence is Arabidopsis thaliana ANT mutant protein. This sequence is
 CC used in the exemplification of the invention.
 XX SQ Sequence 275 AA;
 Query Match 0.8%; Score 8; DB 7; Length 275;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 LKTVALTL 12
 DB 256 LKTVALTL 263
 RESULT 56
 ADG88440
 ID ADG88440 standard; protein; 277 AA.
 XX AC ADG88440;
 XX DT 11-MAR-2004 (first entry)
 XX DE Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANTfDN3.
 XX KW Plant; genetic engineering; cell proliferation; aintegumenta; ANT;
 KW transgenic; transgenic plant; organ mass alteration; fertility;
 KW asexual reproduction; mutant; muten.
 XX OS Synthetic.
 OS Arabidopsis thaliana.
 XX PN US2003159180-A1.
 XX PD 21-AUG-2003.
 XX PF 28-JAN-2002; 2002US-00059911.
 XX PR 28-JAN-2002; 2002US-00059911.
 XX PA (REGC) UNIV CALIFORNIA OFFICE TECHNOLOGY.
 XX

CC particularly relates to a method of increasing cell proliferation in a
 CC plant. The method involves introducing into the plant an expression
 CC cassette containing a plant promoter operably linked to a nucleic acid
 CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2
 CC domain and selecting plants the with increased size or mass. The
 CC invention is useful to generate transgenic plants. The method is useful
 CC for increasing cell proliferation in a plant for altering organ mass,
 CC controlling fertility or enhancing asexual reproduction. The present
 CC sequence is Arabidopsis thaliana ANT mutant protein. This sequence is
 CC used in the exemplification of the invention.

XX Sequence 308 AA;

Query Match 0.8%; Score 8; DB 7; Length 308;

Best Local Similarity 100.0%; Pred. No. 1.6e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0;

QY 5 LKTVALTL 12

|||||

Db 289 LKTVALTL 296

RESULT 59

AAV03187

ID AAY03187 standard; protein; 319 AA.

XX AC AAY03187;

XX DT 16-JUN-1999 (first entry)

XX DE Rat Acid sensitive ion channel protein sequence.

XX KW Acid sensitive ion channel; rat; ASIC; pH mediated pain disorder;

XX KW ischaemia; Gene therapy; proton-gated ion channel.

XX OS Rattus sp.

XX PN WO9911784-A1.

XX PD 11-MAR-1999.

XX PF 28-AUG-1998; 98WO-GB002609.

XX PR 29-AUG-1997; 97GB-00018365.

XX PA (UNLO) UNIV COLLEGE LONDON.

XX PI Wood JN, England S, Akopian AN, Chen C;

XX DR WPI; 1999-205188/17.

XX DR N-PSDB; AAX28162.

XX AC Acid sensitive ion channel (ASIC) proteins - useful in gene therapy for
 treatment of pH mediated pain disorders.

XX PS Claim 5; Page 51-53; 62pp; English.

XX CC This sequence represents an acid sensitive ion channel (ASIC) of the
 invention, isolated from rat dorsal root ganglion. Antisense ASIC DNA or
 RNA is useful in gene therapy for downgrading expression of ASIC protein,
 CC for pH mediated pain disorders e.g. in ischaemia. The vectors are useful
 CC for incorporating ASIC DNA or RNA for use in gene therapy. Proton-gated
 CC ion channel agonists and antagonists are identified using cells
 CC transfected with ASIC DNA by allowing interaction between the candidate
 CC substance and ASIC protein in the membrane, and measuring interaction
 CC and/or cell response. Partial agonists and antagonists can be identified
 CC by their ability to block the response of the cell to present in a
 CC solution of a given acid pH or any agonist. The hybridisation probes are
 CC useful for screening libraries for ASIC DNA or RNA

XX SQ Sequence 319 AA;

Query Match 0.8%; Score 8; DB 2; Length 319;

Best Local Similarity 100.0%; Pred. No. 1.6e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0;

QY 5 LKTVALTL 12

|||||

Db 289 LKTVALTL 296

RESULT 59

AAV03187

ID AAY03187 standard; protein; 319 AA.

XX AC AAY03187;

XX DT 16-JUN-1999 (first entry)

XX DE Rat Acid sensitive ion channel protein sequence.

XX KW Acid sensitive ion channel; rat; ASIC; pH mediated pain disorder;

XX KW ischaemia; Gene therapy; proton-gated ion channel.

XX OS Rattus sp.

XX PN WO9911784-A1.

XX PD 11-MAR-1999.

XX PF 28-AUG-1998; 98WO-GB002609.

XX PR 29-AUG-1997; 97GB-00018365.

XX PA (UNLO) UNIV COLLEGE LONDON.

XX PI Wood JN, England S, Akopian AN, Chen C;

XX DR WPI; 1999-205188/17.

XX DR N-PSDB; AAX28162.

XX AC Acid sensitive ion channel (ASIC) proteins - useful in gene therapy for
 treatment of pH mediated pain disorders.

XX PS Claim 5; Page 51-53; 62pp; English.

XX CC This sequence represents an acid sensitive ion channel (ASIC) of the
 invention, isolated from rat dorsal root ganglion. Antisense ASIC DNA or
 RNA is useful in gene therapy for downgrading expression of ASIC protein,
 CC for pH mediated pain disorders e.g. in ischaemia. The vectors are useful
 CC for incorporating ASIC DNA or RNA for use in gene therapy. Proton-gated
 CC ion channel agonists and antagonists are identified using cells
 CC transfected with ASIC DNA by allowing interaction between the candidate
 CC substance and ASIC protein in the membrane, and measuring interaction
 CC and/or cell response. Partial agonists and antagonists can be identified
 CC by their ability to block the response of the cell to present in a
 CC solution of a given acid pH or any agonist. The hybridisation probes are
 CC useful for screening libraries for ASIC DNA or RNA

XX SQ Sequence 319 AA;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 QASSSIG 44

|||||

Db 306 QASSSIG 313

RESULT 60

AAV22200

ID AAY22200 standard; protein; 350 AA.

XX AC AAY22200;

XX DT 10-SEP-1999 (first entry)

XX DE S. frugiperda retinol dehydratase protein sequence.

XX KW Retinol dehydratase; enzyme; cell growth; cell proliferation; retinol;

XX KW 14-hydroxy-4,14-retro-retinol; 14-HRR; immune system; allergic reaction;

XX KW humoral immune response; graft versus host disease; psoriasis; therapy;

XX KW autoimmune disease.

XX OS Spodoptera frugiperda.

XX PN US5928931-A.

XX PD 27-JUL-1999.

XX PF 07-MAY-1997; 97US-00852481.

XX PR 09-MAY-1996; 96US-0017178P.

XX PA (CORR) CORNELL RES FOUND INC.

XX PA (SLOK) SLOAN KETTERING INST CANCER RES.

XX PI Hammerling U, Gruen F, Buck J;

XX DR WPI; 1999-429507/36.

XX DR N-PSDB; AAX84491.

XX PT Nucleic acids encoding retinol dehydratase useful for antagonizing cell
 growth and proliferation mediated by 14-hydroxy-4,14-retro-retinol or
 PT retinol.

XX PS Claim 2; Fig 4; 23pp; English.

XX CC This sequence is the Spodoptera frugiperda retinol dehydratase of the
 invention. The isolated nucleic acids may be used to transform host cells
 CC for the recombinant production of retinol dehydratase enzymes. These
 CC enzymes may be used to antagonise cell growth and proliferation mediated
 CC by retinol and 14-hydroxy-4,14-retro-retinol (14-HRR) and prevent
 CC uncontrolled cellular replication. Processes as diverse as reproduction,
 CC growth, vision and pattern formation and differentiation during
 CC embryogenesis are influenced by retinol and 14-HRR, and so may be
 CC antagonised in this way. In particular, the immune system is influenced
 CC by 14-HRR and so it may be modulated using retinol dehydratase enzymes.
 CC Diseases characterised by humoral immune responses which may be treated
 CC using the enzyme include graft versus host disease, psoriasis, allergic
 CC reactions and autoimmune diseases. The enzyme may be recombinantly
 CC produced both in fermentation culture and in vivo by transformation of
 CC affected cells with the nucleic acids. The nucleic acid molecules may
 CC also be used as probes in assays to detect and quantify expression of
 CC retinol dehydratase

XX SQ Sequence 350 AA;

Query Match 0.8%; Score 8; DB 2; Length 350;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 668 LADGEHFI 675

|||||


```
Db      303 LADGEHFI 310
RESULT 61
ABU17197
ID      ABU17197 standard; protein; 356 AA.
XX
AC      ABU17197;
XX
XX      19-JUN-2003 (first entry)
XX
XX      Protein encoded by Prokaryotic essential gene #2724.
XX
KW      Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX      Acinetobacter baumannii.
XX
XX      WO200277183-A2.
XX
XX      03-OCT-2002.
XX
XX      21-MAR-2002; 2002WO-US009107.
XX
XX      21-MAR-2001; 2001US-00815242.
XX
XX      06-SEP-2001; 2001US-00948993.
XX
XX      25-OCT-2001; 2001US-0342923P.
XX
XX      08-FEB-2002; 2002US-00072851.
XX
XX      06-MAR-2002; 2002US-0362699P.
XX
XX      (ELIT-) ELITRA PHARM INC.
XX
XX      Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX      Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX      WPI; 2003-029926/02.
XX      N-PSDB; ACA21067.
XX
XX      New antisense nucleic acids, useful for identifying proteins or screening
XX      PT for homologous nucleic acids required for cellular proliferation to
XX      PT isolate candidate molecules for rational drug discovery programs.
XX
XX      Claim 25; SEQ ID NO 45121; 1766pp; English.
XX
XX      The invention relates to an isolated nucleic acid comprising any one of
XX      CC the 6213 antisense sequences given in the specification where expression
XX      CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX      CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX      CC encoding a polypeptide whose expression is inhibited by the antisense
XX      CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX      CC polypeptide or its fragment whose expression is inhibited by the
XX      CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX      CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX      CC proliferation or the activity of a gene in an operon required for
XX      CC proliferation; (7) identifying a compound that influences the activity of
XX      CC the gene product or that has an activity against a biological pathway
XX      CC required for proliferation, or that inhibits cellular proliferation; (8)
XX      CC identifying a gene required for cellular proliferation or the biological
XX      CC pathway in which a proliferation-required gene or its gene product lies
XX      CC or a gene on which the test compound that inhibits proliferation of an
XX      CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX      CC compound's activity; (11) a culture comprising strains in which the gene
XX      CC product is overexpressed or underexpressed; (12) determining the extent
XX      CC to which each of the strains is present in a culture or collection of
XX      CC strains; or (13) identifying the target of a compound that inhibits the
XX      CC proliferation of an organism. The antisense nucleic acids are useful for
XX      CC identifying proteins or screening for homologous nucleic acids required
XX      CC for cellular proliferation to isolate candidate molecules for rational
XX      CC drug discovery programs, or for screening homologous nucleic acids
XX      CC required for proliferation in cells other than S. aureus, S. typhimurium,
XX      CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX      CC the target prokaryotic essential genes. Note: The sequence data for this
XX      CC patent did not form part of the printed specification, but was obtained
XX      in electronic format directly from WIPO at

CC      ftp.wipo.int/pub/published_pct_sequences
XX      SQ      Sequence 356 AA;
XX
XX      Query Match      0.8%; Score 8; DB 6; Length 356;
XX      Best Local Similarity 100.0%; Pred. No. 1.8e+02;
XX      Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
XX
XX      QY      997 LSLITAIK 1004
XX      Db      156 LSLITAIK 163
XX
XX      RESULT 62
XX      ADA33080
XX      ID      ADA33080 standard; protein; 364 AA.
XX
XX      AC      ADA33080;
XX
XX      DT      20-NOV-2003 (first entry)
XX
XX      DE      Acinetobacter baumannii protein #241.
XX
XX      KW      Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX      plant biocontrol agent.
XX
XX      OS      Acinetobacter baumannii.
XX
XX      PN      US6562958-B1.
XX
XX      PD      13-MAY-2003.
XX
XX      PF      04-JUN-1999; 99US-00328352.
XX
XX      PR      09-JUN-1998; 98US-0088701P.
XX
XX      PA      (GENO-) GENOME THERAPEUTICS CORP.
XX
XX      PI      Breton G, Bush D;
XX
XX      DR      WPI; 2003-576092/54.
XX      N-PSDB; ADA28954.
XX
XX      PT      New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
XX      PT for diagnosing a bacterial disease, as components of antibacterial
XX      PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
XX      PT plants.
XX
XX      PS      Example; SEQ ID NO 4367; 328pp; English.
XX
XX      CC      The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX      CC The A. baumannii nucleic acids and polypeptides are useful as reagents
XX      CC for diagnosing a bacterial disease, as components of antibacterial
XX      CC vaccines, as targets for antibacterial drugs, to detect the presence of
XX      CC A. baumannii and other Acinetobacter species in a sample, in screening
XX      CC compounds for the ability to interfere with the A. baumannii life cycle
XX      CC or to inhibit A. baumannii infection, and as biocontrol agents for
XX      CC plants. The present sequence represents the amino acid sequence of an A.
XX      CC baumannii protein.
XX
XX      SQ      Sequence 364 AA;
XX
XX      Query Match      0.8%; Score 8; DB 6; Length 364;
XX      Best Local Similarity 100.0%; Pred. No. 1.9e+02;
XX      Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
XX
XX      QY      997 LSLITAIK 1004
XX      Db      164 LSLITAIK 171
XX
XX      RESULT 63
XX      ABP73447
```

ID ABP73447 standard; protein; 411 AA.
XX
AC ABP73447;
XX
DT 30-JAN-2003 (first entry)
XX
XX Candida albicans essential protein SEQ ID NO 7284.
XX
DE
XX Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
XX
OS Candida albicans.
XX
XX WO200253728-A2.
PN
XX 11-JUL-2002.
PD
XX 26-DEC-2001; 2001WO-US049486.
PF
XX 29-DEC-2000; 2000US-0259128P.
PR
XX 20-FEB-2001; 2001US-00792024.
PR
XX 22-AUG-2001; 2001US-0314050P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KU;
PI
XX WPI; 2002-566694/60.
DR
XX N-PSDB; AB231997.
DR
XX Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
PT
XX
XX Claim 44; SEQ ID NO 7284; 167pp + Sequence Listing; English.
XX
XX The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance and/or pathogenicity of a diploid fungus
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX
SQ Sequence 411 AA;

Query Match 0.8%; Score 8; DB 5; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 795 LEKELDLL 802
|||||
DB 37 LEKELDLL 44

RESULT 64
ADS28368
ID ADS28368 standard; protein; 458 AA.
XX
AC ADS28368;
XX
DT 02-DEC-2004 (first entry)
XX
XX Bacterial polypeptide #17401.
DE
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
XX US2003233675-A1.
PN
XX 18-DEC-2003.
PD
XX 20-FEB-2003; 2003US-00369493.
PF
XX 21-FEB-2002; 2002US-0360039P.
PR
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
PI
XX WPI; 2004-061375/06.
DR
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
PT
XX Claim 1; SEQ ID NO 17401; 122pp; English.
PS
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 458 AA;

Query Match 0.8%; Score 8; DB 8; Length 458;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 KATAKAEK 757
Db 447 KATAKAEK 454
|||||

RESULT 65
ADG88438
ID ADG88438 standard; protein; 473 AA.
XX
AC ADG88438;
XX
XX 11-MAR-2004 (first entry)
XX
XX Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANTND1.
XX
XX Plant; genetic engineering; cell proliferation; aintegumenta; ANT;
KW transgenic; transgenic plant; organ mass alteration; fertility;
KW asexual reproduction; mutant; mutein.
XX
XX Synthetic.
OS Arabidopsis thaliana.
XX
XX US2003159180-A1.
XX
XX 21-AUG-2003.
XX
XX 28-JAN-2002; 2002US-00059911.
PF
XX 28-JAN-2002; 2002US-00059911.
PR
XX
XX (REGC) UNIV CALIFORNIA OFFICE TECHNOLOGY.
XX
XX Fischer RL, Mizukami Y;
PI N-PSDB; ADG88455.
DR
XX
XX Increasing cell proliferation in a plant by introducing into the plant an
PT expression cassette containing a plant promoter operably linked to a
PT nucleic acid encoding a modified ANT polypeptide and selecting plants
PT with increased mass.
XX
XX Claim 12; SEQ ID NO 9; Sipp; English.
PS

CC The present invention relates to plant genetic engineering. The invention
CC particularly relates to a method of increasing cell proliferation in a
CC plant. The method involves introducing into the plant an expression
CC cassette containing a plant promoter operably linked to a nucleic acid
CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2
CC domain and selecting plants the with increased size or mass. The
CC invention is useful to generate transgenic plants. The method is useful
CC for increasing cell proliferation in a plant for altering organ mass,
CC controlling fertility or enhancing asexual reproduction. The present
CC sequence is Arabidopsis thaliana ANT mutant protein. This sequence is
XX used in the exemplification of the invention.
SQ Sequence 473 AA;

Query Match 0.8%; Score 8; DB 7; Length 473;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALT 12
Db 454 LKTVALT 461
|||||

RESULT 66
ADC08227
ID ADC08227 standard; protein; 480 AA.

XX ADC08227;
AC 18-DEC-2003 (first entry)
DT Rice protein sequence Seq ID532 related to grain filling.
XX
XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW gene; ds; plant.
XX
XX Oryza sativa.
OS
XX WO2003009095-A2.
PN
XX 03-JAN-2003.
PD
XX 21-JUN-2002; 2002WO-IB002450.
PF
XX 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0325277P.
PR 20-DEC-2001; 2001US-0342327P.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
PA
XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;
PI
XX WPI; 2003-229341/22.
DR N-PSDB; ADC08226.
XX
XX New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.
XX
XX Claim 34; SEQ ID NO 532; 130pp; English.
PS
XX This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is the
CC amino acid sequence of a rice protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/publishedpct_sequences.
XX
XX Sequence 480 AA;
SQ

Query Match 0.8%; Score 8; DB 7; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 55 GLVEGKGP 62
|||||

RESULT 67
AAB92603
ID AAB92603 standard; protein; 483 AA.
XX
AC AAB92603;
XX
XX 26-JUN-2001 (first entry)
DT

XX Human protein sequence SEQ ID NO:10860.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-00116136.
 XX 29-JUL-1999; 937P-00248036.
 XX 27-AUG-1999; 950P-00300253.
 XX 11-JAN-2000; 2000JP-00118776.
 XX 02-MAY-2000; 2000JP-00183767.
 XX 09-JUN-2000; 2000JP-00241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 XX Claim 8; SEQ ID NO 10860; 2537pp + Sequence Listing; English.
 XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 XX Sequence 483 AA;
 Query Match 0.8%; Score 8; DB 4; Length 483;
 Best Local Similarity 100.0%; Pred.No. 2.4e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 945 EKSSSES 952
 Db 444 EKSSSES 451
 RESULT 68
 ADJ48708
 ID ADJ48708 standard; protein; 488 AA.
 XX
 XX AC ADJ48708;
 XX

DT 06-MAY-2004 (first entry)
 XX Oil-associated gene related protein #208.
 DE Oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
 XX Unidentified.
 OS US2004025202-A1.
 PN 05-FEB-2004.
 XX 14-MAR-2003; 2003US-00389566.
 XX 15-MAR-2002; 2002US-0385301P.
 PR 26-JUN-2002; 2002US-0391786P.
 PR 26-JUN-2002; 2002US-0392018P.
 XX (LAURIE) LAURIE C C.
 PA (RAVA) RAVANELLO M.
 PA (SAVA) SAVAGE T.
 PA (LEDE) LEDEAUX J R.
 XX (ROGE) ROGERS J A.
 PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
 XX WPI; 2004-142683/14.
 DR Novel recombinant DNA construct comprising a promoter functional in plants operably linked to an oil-associated gene for producing transgenic plant seed.
 PT Example 3; SEQ ID NO 712; 22pp; English.
 XX The invention relates to a recombinant DNA construct comprising a promoter functional in plants operably linked to an oil-associated gene. The construct is useful for transgenic plant seed which has in its genome the construct, that is functional in the plant to transcribe the oil-associated gene. The transgenic plant seed grows into a plant having enhanced seed oil as compared to wild type. The construct is useful for producing hybrid maize seed. The transgenic plant seed is useful for producing vegetable oil. The present sequence represents the amino acid sequence of an oil-associated gene related protein.
 XX Sequence 488 AA;
 Query Match 0.8%; Score 8; DB 8; Length 488;
 Best Local Similarity 100.0%; Pred.No. 2.4e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 804 GLVGGKGP 811
 Db 55 GLVGGKGP 62
 RESULT 69
 ADJ49363
 ID ADJ49363 standard; protein; 488 AA.
 XX
 XX AC ADJ49363;
 XX 06-MAY-2004 (first entry)
 DT Oil-associated gene related protein #863.
 DE oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
 XX Unidentified.
 OS US2004025202-A1.
 PN 05-FEB-2004.
 XX

CC The construct is useful for transgenic plant seed which has in its genome
 CC the construct, that is functional in the plant to transcribe the oil-
 CC associated gene. The transgenic plant seed grows into a plant having
 CC enhanced seed oil as compared to wild type. The construct is useful for
 CC producing hybrid maize seed. The transgenic plant seed is useful for
 CC producing vegetable oil. The present sequence represents the amino acid
 CC sequence of an oil-associated gene related protein.

XX Sequence 488 AA;

Query Match 0.8%; Score 8; DB 8; Length 488;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
 |||||
 Db 55 GLVEGKGP 62

RESULT 72

ADC07838
 ID ADC07838 standard; protein; 496 AA.

AC ADC07838;

DT 18-DEC-2003 (first entry)

DE Rice protein sequence Seq ID104 related to grain filling.

XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
 KW Carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
 KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
 KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
 KW Gene; ds; plant.

XX Oryza sativa.

XX WO200300905-A2.

XX 03-JAN-2003.

XX 21-JUN-2002; 2002WO-IB002450.

XX 22-JUN-2001; 2001US-0300112P.

XX 26-SEP-2001; 2001US-0325277P.

XX 20-DEC-2001; 2001US-0342327P.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
 PI Glazebrook J, Katagiri F, Kreps J, Provart N, Rickes D;

XX WPI; 2003-229341/22.

DR N-PSDB; ADC07837.

XX New plant genes encoding polypeptides having an activity involved in or
 PT associated with the synthesis, metabolism or degradation of carbohydrates
 PT in the plant grain useful in generating plants having improved
 PT nutritional properties.

XX Claim 1; SEQ ID NO 104; 130pp; English.

XX This invention, in the area of plant biotechnology, relates to novel
 CC polynucleotides comprising a nucleotide sequence encoding a protein which
 CC is involved in or associated with the synthesis, metabolism or
 CC degradation of carbohydrates in the plant grain and the expression of
 CC which is up-regulated during grain filling. The plant is selected from
 CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
 CC sugarbeet, wheat, and rice. The invention may be useful for the
 CC improvement of protein, oil, starch, fibre and moisture content of the
 CC cereal grains. In addition, carbohydrate levels may be modified to a more
 CC desirable level using the present invention. The present sequence is the
 CC amino acid sequence of a rice protein of the invention. Note: The

CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/publishedpct_sequences.

SQ Sequence 496 AA;

Query Match 0.8%; Score 8; DB 7; Length 496;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
 |||||
 Db 55 GLVEGKGP 62

RESULT 73

ADJ50282
 ID ADJ50282 standard; protein; 500 AA.

XX ADJ50282;

XX 06-MAY-2004 (first entry)

DE Oil-associated gene related protein #1782.

XX oil-associated gene; transgenic; enhanced seed oil; vegetable oil.

XX Unidentified.

XX US2004025202-A1.

XX 05-FEB-2004.

XX 14-MAR-2003; 2003US-00389566.

XX 15-MAR-2002; 2002US-0365301P.

XX 26-JUN-2002; 2002US-0391786P.

XX 26-JUN-2002; 2002US-0392018P.

XX (LAUR/) LAURIE C C.

XX (RAVA/) RAVANELLO M.

XX (SAVA/) SAVAGE T.

XX (LEDE/) LEDEUX J R.

XX (ROGE/) ROGERS J A.

XX Laurie CC, Ravanello M, Savage T, Ledoux JR, Rogers JA;

XX WPI; 2004-142683/14.

XX Novel recombinant DNA construct comprising a promoter functional in
 PT plants operably linked to an oil-associated gene for producing transgenic
 PT plant seed.

XX Example 3; SEQ ID NO 2286; 22pp; English.

XX The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in plants operably linked to an oil-associated gene.
 CC The construct is useful for transgenic plant seed which has in its genome
 CC the construct, that is functional in the plant to transcribe the oil-
 CC associated gene. The transgenic plant seed grows into a plant having
 CC enhanced seed oil as compared to wild type. The construct is useful for
 CC producing hybrid maize seed. The transgenic plant seed is useful for
 CC producing vegetable oil. The present sequence represents the amino acid
 CC sequence of an oil-associated gene related protein.

SQ Sequence 500 AA;

Query Match 0.8%; Score 8; DB 8; Length 500;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
 |||||

```

Db      51 GLVEGKGP 58

RESULT 74
ADJ49407
ID      ADJ49407 standard; protein; 503 AA.
AC      ADJ49407;
XX
XX      06-MAY-2004 (first entry)
XX
XX      Oil-associated gene related protein #907.
XX
XX      oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX
XX      Unidentified.
XX
XX      US2004025202-A1.
XX
XX      05-FEB-2004.
XX
XX      14-MAR-2003; 2003US-00389566.
XX
XX      15-MAR-2002; 2002US-0365301P.
XX      26-JUN-2002; 2002US-0391786P.
XX      26-JUN-2002; 2002US-0392018P.
XX
XX      (LAUR/) LAURIE C C.
XX      (SAVA/) SAVAGE T.
XX      (LEDE/) LEDEAUX J R.
XX      (ROGE/) ROGERS J A.
XX
XX      Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX      WPI; 2004-142683/14.
XX
XX      Novel recombinant DNA construct comprising a promoter functional in
XX      plants operably linked to an oil-associated gene for producing transgenic
XX      plant seed.
XX
XX      Example 3; SEQ ID NO 1411; 22pp; English.
XX
XX      The invention relates to a recombinant DNA construct comprising a
XX      promoter functional in plants operably linked to an oil-associated gene.
XX      The construct is useful for transgenic plant seed which has in its genome
XX      the construct, that is functional in the plant to transcribe the oil-
XX      associated gene. The transgenic plant seed grows into a plant having
XX      enhanced seed oil as compared to wild type. The construct is useful for
XX      producing hybrid maize seed. The transgenic plant seed is useful for
XX      producing vegetable oil. The present sequence represents the amino acid
XX      sequence of an oil-associated gene related protein.
XX
XX      Sequence 503 AA;

Query Match      0.8%; Score 8; DB 8; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      804 GLVEGKGP 811
        |||||
Db      55 GLVEGKGP 62

RESULT 75
ADJ48856
ID      ADJ48856 standard; protein; 503 AA.
XX
XX      ADJ48856;
XX
XX      06-MAY-2004 (first entry)
XX
XX      Oil-associated gene related protein #356.

```

```

XX      oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX
XX      Unidentified.
XX
XX      US2004025202-A1.
XX
XX      05-FEB-2004.
XX
XX      14-MAR-2003; 2003US-00389566.
XX
XX      15-MAR-2002; 2002US-0365301P.
XX      26-JUN-2002; 2002US-0391786P.
XX      26-JUN-2002; 2002US-0392018P.
XX
XX      (LAUR/) LAURIE C C.
XX      (SAVA/) SAVAGE T.
XX      (LEDE/) LEDEAUX J R.
XX      (ROGE/) ROGERS J A.
XX
XX      Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX      WPI; 2004-142683/14.
XX
XX      Novel recombinant DNA construct comprising a promoter functional in
XX      plants operably linked to an oil-associated gene for producing transgenic
XX      plant seed.
XX
XX      Example 3; SEQ ID NO 860; 22pp; English.
XX
XX      The invention relates to a recombinant DNA construct comprising a
XX      promoter functional in plants operably linked to an oil-associated gene.
XX      The construct is useful for transgenic plant seed which has in its genome
XX      the construct, that is functional in the plant to transcribe the oil-
XX      associated gene. The transgenic plant seed grows into a plant having
XX      enhanced seed oil as compared to wild type. The construct is useful for
XX      producing hybrid maize seed. The transgenic plant seed is useful for
XX      producing vegetable oil. The present sequence represents the amino acid
XX      sequence of an oil-associated gene related protein.
XX
XX      Sequence 503 AA;

Query Match      0.8%; Score 8; DB 8; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      804 GLVEGKGP 811
        |||||
Db      55 GLVEGKGP 62

RESULT 76
ADJ50184
ID      ADJ50184 standard; protein; 503 AA.
XX
XX      ADJ50184;
XX
XX      06-MAY-2004 (first entry)
XX
XX      Oil-associated gene related protein #1684.
XX
XX      oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX
XX      Unidentified.
XX
XX      US2004025202-A1.
XX
XX      05-FEB-2004.
XX
XX      14-MAR-2003; 2003US-00389566.
XX      15-MAR-2002; 2002US-0365301P.

```



```

XX SQ Sequence 523 AA;
Query Match 0.8%; Score 8; DB 2; Length 523;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 51 GLVEGKGP 58

RESULT 80
ADJ48855
ID ADJ48855 standard; protein; 527 AA.
XX
AC ADJ48855;
XX
DT 06-MAY-2004 (first entry)
XX
DE Oil-associated gene related protein #355.
XX
KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX
OS Unidentified.
XX
PN US2004025202-A1.
XX
PD 05-FEB-2004.
XX
PF 14-MAR-2003; 2003US-00389566.
XX
PR 15-MAR-2002; 2002US-0365301P.
XX
PR 26-JUN-2002; 2002US-0391786P.
XX
PR 26-JUN-2002; 2002US-0392018P.
XX
(LAUR/) LAURIE C C.
PA (RAVA/) RAVANELLO M.
PA (SAVA/) SAVAGE T.
PA (LEDE/) LEDEAUX J R.
PA (ROGE/) ROGERS J A.
XX
PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX
WPI; 2004-142683/14.
XX
PT Novel recombinant DNA construct comprising a promoter functional in
PT plants operably linked to an oil-associated gene for producing transgenic
PT plant seed.
XX
PS Example 3; SEQ ID NO 859; 22pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in plants operably linked to an oil-associated gene.
CC The construct is useful for transgenic plant seed which has in its genome
CC the construct, that is functional in the plant to transcribe the oil-
CC associated gene. The transgenic plant seed grows into a plant having
CC enhanced seed oil as compared to wild type. The construct is useful for
CC producing hybrid maize seed. The transgenic plant seed is useful for
CC producing vegetable oil. The present sequence represents the amino acid
CC sequence of an oil-associated gene related protein.
XX
SQ Sequence 527 AA;
Query Match 0.8%; Score 8; DB 8; Length 527;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 57 GLVEGKGP 64

RESULT 81
AAG31382
ID AAG31382 standard; protein; 528 AA.
XX
AC AAG31382;

XX SQ Sequence 526 AA;
Query Match 0.8%; Score 8; DB 2; Length 526;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;

XX SQ Sequence 523 AA;
Query Match 0.8%; Score 8; DB 2; Length 523;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 43 GLVEGKGP 50

RESULT 79
AAR88327
ID AAR88327 standard; protein; 526 AA.
XX
AC AAR88327;
XX
DT 08-AUG-1996 (first entry)
XX
DE Beta-amylase S291A, S346P mutein.
XX
KW Beta-amylase; 1,4-alpha-D-glucan maltohydrolase; EC 3.2.1.2;
KW thermostable; alkaline stable; mutant; mutein; variant; barley.
XX
OS Hordeum vulgare.
XX
FH Key Location/Qualifiers
FT Misc-difference 291 /label= S291A
FT /note= "wild-type Ser has been replaced by Ala"
FT Misc-difference 346
FT /label= S346P
FT /note= "wild-type Ser has been replaced by Pro"
FT Misc-difference 472..473
FT /note= "The corresp. coding sequence (see AAT03695) codes
FT for amino acid residues Pro-Phe-Pro- Phe-Gln between
FT Gln472 and Glu473 of the present sequence to give a
FT protein that is 531 amino acids long"
XX
JP07327681-A.
XX
PD 19-DEC-1995.
XX
PF 08-JUN-1994; 94JP-00126151.
XX
PR 08-JUN-1994; 94JP-00126151.
XX
PA (SAPB ) SAPPORO BREWERIES.
XX
WPI; 1996-072339/08.
XX
N-PSDB; AAT03695.
XX
PT Recombinant beta-amylase variant - has improved heat stability and
PT improved stability in alkaline conditions.
XX
PS Claim 1; Page 5-6; 11pp; Japanese.
XX
CC The present sequence is that of a beta-amylase variant in which wild-
CC type Ser residues at positions 291 and 346 have been replaced by an Ala
CC and a Pro residue, respectively. The resulting mutein has improved heat
CC stability, optimum pH range of 3.5-6.5 and stable pH region of 3.5-12.5
CC (i.e. the enzyme shows residual activity of at least 80% in this range).
CC The enzyme removes beta-maltose units from the non-reducing terminal of
CC polysaccharides (esp. starch, amylose and a maltooligosaccharide with a
CC degree of polymerisation of 3 or more). N.B. In the SEQ.ID. listing, the
CC beta-amylase sequence is described as being 531 amino acid residues long
CC but only 526 residues are given in the specification
XX
SQ Sequence 526 AA;
Query Match 0.8%; Score 8; DB 2; Length 526;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;

```



```

PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      0.8%; Score 8; DB 3; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 LKTVALT 12
Db      509 LKTVALT 516

RESULT 82
AAG31381
ID AAG31381 standard; protein; 529 AA.
XX
AC AAG31381;
XX
DT
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37676.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
```

termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

99US-0121825P.

99US-0123180P.

99US-0123548P.

99US-0125788P.

99US-0126264P.

99US-0126785P.

99US-0127462P.

99US-0128234P.

99US-0128714P.

99US-0129845P.

99US-0130077P.

99US-0130449P.

99US-0130510P.

99US-0130891P.

99US-0131449P.

99US-0132407P.

99US-0132484P.

99US-0132485P.

99US-0132486P.

99US-0132863P.

99US-0132867P.

99US-0134256P.

99US-0134218P.

99US-0134219P.

99US-0134221P.

99US-0134370P.

99US-0134768P.

99US-0134941P.

99US-0135124P.

99US-0135353P.

99US-0135629P.

99US-0136021P.

99US-0136392P.

99US-0136782P.

99US-0137222P.

99US-0137528P.

99US-0139119P.

99US-0139452P.

99US-0139453P.

99US-0139492P.

99US-0139454P.

99US-0139455P.

99US-0139456P.

99US-0139457P.

99US-0139458P.

99US-0139459P.

99US-0139460P.

99US-0139461P.

99US-0139462P.

99US-0139463P.

99US-0139750P.

99US-0139763P.

99US-0139817P.

99US-0139899P.

99US-0140353P.

99US-0140354P.

99US-0140695P.

PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 27-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.

PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.8%; Score 8; DB 3; Length 529;
Best Local Similarity 100.0%; Pred.No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTL 12
| | | | |
Db 510 LKTVALTL 517

RESULT 83

AAR67005
ID AAR67005 standard; protein; 530 AA.
XX
AC AAR67005;
XX
DT 17-AUG-1995 (first entry)
XX
DE Barley beta-amylase.
XX
KW Barley beta-amylase; recombinant production.
XX
OS Hordeum vulgare.
XX
PN JP06303983-A.
XX
PD 01-NOV-1994.
XX
PF 18-FEB-1994; 94JP-00058119.

```
XX 26-FEB-1993; 93JP-00038870.
XX (SAPB ) SAPPORO BREWERIES.
XX WPI; 1995-018276/03.
XX DR N-PSDB; AAQ74813.
XX
XX Barley beta-amylase gene - for the efficient production of beta-amylase.
XX
XX Claim 1; Page 12-13; 19pp; Japanese.
XX
XX AAQ74813 encodes AAR67005 barley beta-amylase. AAQ74813 can be used in
XX the construction of a vector, for the recombinant expression of barley
XX beta-amylase
XX
XX Sequence 530 AA;
SQ
Query Match 0.8%; Score 8; DB 2; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62
|||||
55 GLVEGKGP 62

RESULT 84
AAR97613
ID AAR97613 standard; protein; 531 AA.
XX AC
XX AAR97613;
XX
XX 20-AUG-1996 (first entry)
XX
XX Beta-amylase sevenfold mutant.
XX
XX Beta-amylase; maltose; thermostable enzyme; protein engineering;
XX enzyme engineering; barley; Hordeum vulgare.
XX
XX Synthetic.
XX
XX EP713916-A2.
XX
XX 29-MAY-1996.
XX
XX 27-SEP-1995; 95EP-00115255.
XX
XX 28-SEP-1994; 94JP-00233086.
XX
XX (SAPB ) SAPPORO BREWERIES.
XX
XX Yoshigi N, Maeba H, Okada Y;
XX
XX WPI; 1996-253043/26.
XX DR N-PSDB; AAT29192.
XX
XX Recombinant beta-amylase used in the prodn. of maltose - has superior
XX thermostability compared to wild type enzymes.
XX
XX Claim 2; Page 18-19; 29pp; English.
XX
XX A 7-fold mutant beta-amylase (AAR97613) has a thermostability improved by
XX 11.6 degC compared with the native (barley) enzyme and 5.8 degC compared
XX with soybean beta-amylase. It contains substitutions of Leu for Met-181,
XX Ala for Ser-291, Val for Ile-293, Pro for Ser-347, Pro for Ser-346, Asp
XX for Gln-348 and Ser for Ala-372. The mutant beta-amylase can be expressed
XX in transformed Escherichia coli JM109 cells carrying a vector (see also
XX AAT29193) that incorporates the mutated coding sequence (AAT29192). It is
XX useful for industrial prodn. of maltose
XX
XX Sequence 531 AA;
SQ
Query Match 0.8%; Score 8; DB 2; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62
|||||
55 GLVEGKGP 62

RESULT 84
AAR97613
ID AAR97613 standard; protein; 531 AA.
XX AC
XX AAR97613;
XX
XX 20-AUG-1996 (first entry)
XX
XX Beta-amylase sevenfold mutant.
XX
XX Beta-amylase; maltose; thermostable enzyme; protein engineering;
XX enzyme engineering; barley; Hordeum vulgare.
XX
XX Synthetic.
XX
XX EP713916-A2.
XX
XX 29-MAY-1996.
XX
XX 27-SEP-1995; 95EP-00115255.
XX
XX 28-SEP-1994; 94JP-00233086.
XX
XX (SAPB ) SAPPORO BREWERIES.
XX
XX Yoshigi N, Maeba H, Okada Y;
XX
XX WPI; 1996-253043/26.
XX DR N-PSDB; AAT29192.
XX
XX Recombinant beta-amylase used in the prodn. of maltose - has superior
XX thermostability compared to wild type enzymes.
XX
XX Claim 2; Page 18-19; 29pp; English.
XX
XX A 7-fold mutant beta-amylase (AAR97613) has a thermostability improved by
XX 11.6 degC compared with the native (barley) enzyme and 5.8 degC compared
XX with soybean beta-amylase. It contains substitutions of Leu for Met-181,
XX Ala for Ser-291, Val for Ile-293, Pro for Ser-347, Pro for Ser-346, Asp
XX for Gln-348 and Ser for Ala-372. The mutant beta-amylase can be expressed
XX in transformed Escherichia coli JM109 cells carrying a vector (see also
XX AAT29193) that incorporates the mutated coding sequence (AAT29192). It is
XX useful for industrial prodn. of maltose
XX
XX Sequence 531 AA;
SQ
Query Match 0.8%; Score 8; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 51 GLVEGKGP 58
|||||
51 GLVEGKGP 58

RESULT 85
ADJ49017
ID ADJ49017 standard; protein; 533 AA.
XX AC
XX ADJ49017;
XX
XX 06-MAY-2004 (first entry)
XX
XX Oil-associated gene related protein #517.
XX
XX oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX
XX Unidentified.
XX
XX US2004025202-A1.
XX
XX 05-FEB-2004.
XX
XX 14-MAR-2003; 2003US-00389566.
XX
XX 15-MAR-2002; 2002US-0365301P.
XX PR 26-JUN-2002; 2002US-0391786P.
XX PR 26-JUN-2002; 2002US-0392018P.
XX
XX (LAUR/) LAURIE C C.
XX PA (RAVA/) RAVANELLO M.
XX PA (SAVA/) SAVAGE T.
XX PA (LEDE/) LEDEAUX J R.
XX PA (ROGE/) ROGERS J A.
XX
XX Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX
XX WPI; 2004-142683/14.
XX
XX Novel recombinant DNA construct comprising a promoter functional in
XX plants operably linked to an oil-associated gene for producing transgenic
XX plant seed.
XX
XX Example 3; SEQ ID NO 1021; 22pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in plants operably linked to an oil-associated gene.
XX The construct is useful for transgenic plant seed which has in its genome
XX the construct, that is functional in the plant to transcribe the oil-
XX associated gene. The transgenic plant seed grows into a plant having
XX enhanced seed oil as compared to wild type. The construct is useful for
XX producing hybrid maize seed. The transgenic plant seed is useful for
XX producing vegetable oil. The present sequence represents the amino acid
XX sequence of an oil-associated gene related protein.
XX
XX Sequence 533 AA;
SQ
Query Match 0.8%; Score 8; DB 8; Length 533;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 53 GLVEGKGP 60
|||||
53 GLVEGKGP 60

RESULT 86
ADJ49016
ID ADJ49016 standard; protein; 533 AA.
```

```

XX AC ADJ49016;
XX DT 06-MAY-2004 (first entry)
XX DE Oil-associated gene related protein #516.
XX KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX OS Unidentified.
XX PN US2004025202-A1.
XX PD 05-FEB-2004.
XX PF 14-MAR-2003; 2003US-00389566.
XX PR 15-MAR-2002; 2002US-0365301P.
XX PR 26-JUN-2002; 2002US-0391786P.
XX PR 26-JUN-2002; 2002US-0392018P.
XX (LAUR/) LAURIE C C.
XX PA (RAVA/) RAVANELLO M.
XX PA (SAVA/) SAVAGE T.
XX PA (LEDE/) LEDEUX J R.
XX PA (ROGE/) ROGERS J A.
XX PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX WI; 2004-142683/14.
XX DR
XX PT Novel recombinant DNA construct comprising a promoter functional in
XX plants operably linked to an oil-associated gene for producing transgenic
XX plant seed.
XX Example 3; SEQ ID NO 1020; 22pp; English.
XX PS
XX CC The invention relates to a recombinant DNA construct comprising a
XX promoter functional in plants operably linked to an oil-associated gene.
XX CC The construct is useful for transgenic plant seed which has in its genome
XX the construct, that is functional in the plant to transcribe the oil-
XX associated gene. The transgenic plant seed grows into a plant having
XX enhanced seed oil as compared to wild type. The construct is useful for
XX producing hybrid maize seed. The transgenic plant seed is useful for
XX producing vegetable oil. The present sequence represents the amino acid
XX sequence of an oil-associated gene related protein.
XX SQ Sequence 533 AA;

Query Match 0.8%; Score 8; DB 8; Length 533;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 53 GLVEGKGP 60

RESULT 87
AAW04261
ID AAW04261 standard; protein; 535 AA.
XX AC AAW04261;
XX DT 16-JUN-1997 (first entry)
XX DE Beta-amylase protein.
XX KW Beta-amylase; barley; structural protein; thermally stable enzyme.
XX OS Hordeum vulgare.
XX PN WO9630525-A1.

```

```

XX PD 03-OCT-1996.
XX PF 27-MAR-1996; 96WO-JP000799.
XX PR 27-MAR-1995; 95JP-00092004.
XX PA (SAPB ) SAPPORO BREWERIES.
XX PI Yoshigi N, Okada Y;
XX WI; 1996-455370/45.
XX DR N-PSDB; AAT33961.
XX PT Barley beta-amylase structural gene - for prodn. of recombinant beta-
XX amylase, and plants with improved characteristics, useful in beverage and
XX food industries.
XX PS Claim 1; Page 8-11; 30pp; Japanese.
XX CC This sequence represents the barley beta-amylase structural protein. The
XX sequence represented by AAW15746 represents a mutated version of this
XX sequence. The DNA encoding this sequence was isolated from barley seed
XX genomic DNA using the primers represented by AAT33963 and AAT33964. The
XX gene encoding this sequence is used for the production of recombinant
XX beta-amylase with good thermal stability. The gene is also used to create
XX transformants of barley and other plants with improved characteristics.
XX CC The transformants can then be used in the alcoholic beverage (e.g. beer
XX and distilled spirits), food and enzyme industries
XX SQ Sequence 535 AA;

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 88
AAW15746
ID AAW15746 standard; protein; 535 AA.
XX AC AAW15746;
XX DT 16-JUN-1997 (first entry)
XX DE Mutated beta-amylase protein.
XX KW Beta-amylase; barley; structural protein; thermally stable enzyme.
XX OS Hordeum vulgare.
XX PI Key Location/Qualifiers
XX FT Misc-difference 185 /label= M185L
XX FT Misc-difference 295 /label= S295A
XX FT Misc-difference 297 /label= I297V
XX FT Misc-difference 350 /label= S350P
XX FT Misc-difference 351 /label= S351P
XX FT Misc-difference 352 /label= Q352D
XX FT Misc-difference 376 /label= A376S
XX PN WO9630525-A1.

```

```

PD 03-OCT-1996.
XX
PF 27-MAR-1996; 96WO-JP000799.
XX
PR 27-MAR-1995; 95JP-00092004.
XX
PA (SAPB ) SAPPORO BREWERIES.
XX
PI Yoshigi N, Okada Y;
XX
DR WPI; 1996-455370/45.
XX
DR N-PSDB; AA133962.
XX
PT Barley beta-amylase structural gene - for prodn. of recombinant beta-
PT amylase, and plants with improved characteristics, useful in beverage and
PT food industries.
XX
PS Claim 2; Page; 30pp; Japanese.
XX
CC This sequence represents a mutated version of the barley beta-amylase
CC structural protein. The sequence represented by AA15745 represents the
CC wild type version of this sequence. The DNA encoding the wild type
CC sequence was isolated from barley seed genomic DNA using the primers
CC represented by AA133963 and AA133964. The gene encoding this sequence is
CC used for the production of recombinant beta-amylase with good thermal
CC stability. The gene is also used to create transformants of barley and
CC other plants with improved characteristics. The transformants can then be
CC used in the alcoholic beverage (e.g. beer and distilled spirits), food
CC and enzyme industries
XX
SQ Sequence 535 AA;
Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62
|||||||
RESULT 89
ADJ49120
ID ADJ49120 standard; protein; 535 AA.
XX
AC ADJ49120;
XX
DT 06-MAY-2004 (first entry)
XX
DE Oil-associated gene related protein #620.
XX
KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX
OS Unidentified.
XX
PN US2004025202-A1.
XX
PD 05-FEB-2004.
XX
PF 14-MAR-2003; 2003US-00389566.
XX
PR 15-MAR-2002; 2002US-0365301P.
XX
PR 26-JUN-2002; 2002US-0391786P.
XX
PR 26-JUN-2002; 2002US-0392018P.
XX
PA (LAUR/) LAURIE C C.
PA (RAVA/) RAVANELLO M.
PA (SAVA/) SAVAGE T.
PA (LEDE/) LEDEAUX J R.
PA (ROGE/) ROGERS J A.
XX
PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX
WPI; 2004-142683/14.
XX
PT Novel recombinant DNA construct comprising a promoter functional in
PT plants operably linked to an oil-associated gene for producing transgenic
PT plant seed.
XX
PS Example 3; SEQ ID NO 2397; 22pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in plants operably linked to an oil-associated gene.
CC The construct is useful for transgenic plant seed which has in its genome
CC the construct, that is functional in the plant to transcribe the oil-
CC associated gene. The transgenic plant seed grows into a plant having
CC enhanced seed oil as compared to wild type. The construct is useful for
CC producing hybrid maize seed. The transgenic plant seed is useful for
CC producing vegetable oil. The present sequence represents the amino acid
CC sequence of an oil-associated gene related protein.
XX
SQ Sequence 535 AA;
Query Match 0.8%; Score 8; DB 8; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62
|||||||
RESULT 90
ADJ50393
ID ADJ50393 standard; protein; 535 AA.
XX
AC ADJ50393;
XX
DT 06-MAY-2004 (first entry)
XX
DE Oil-associated gene related protein #1893.
XX
KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX
OS Unidentified.
XX
PN US2004025202-A1.
XX
PD 05-FEB-2004.
XX
PF 14-MAR-2003; 2003US-00389566.
XX
PR 15-MAR-2002; 2002US-0365301P.
XX
PR 26-JUN-2002; 2002US-0391786P.
XX
PR 26-JUN-2002; 2002US-0392018P.
XX
PA (LAUR/) LAURIE C C.
PA (RAVA/) RAVANELLO M.
PA (SAVA/) SAVAGE T.
PA (LEDE/) LEDEAUX J R.
PA (ROGE/) ROGERS J A.
XX
PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX
WPI; 2004-142683/14.
XX
PT Novel recombinant DNA construct comprising a promoter functional in
PT plants operably linked to an oil-associated gene for producing transgenic
PT plant seed.
XX
PS Example 3; SEQ ID NO 2397; 22pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in plants operably linked to an oil-associated gene.
CC The construct is useful for transgenic plant seed which has in its genome
CC the construct, that is functional in the plant to transcribe the oil-

```

CC associated gene. The transgenic plant seed grows into a plant having
 CC enhanced seed oil as compared to wild type. The construct is useful for
 CC producing hybrid maize seed. The transgenic plant seed is useful for
 CC producing vegetable oil. The present sequence represents the amino acid
 CC sequence of an oil-associated gene related protein.
 XX
 XX

SQ Sequence 535 AA;

Query Match 0.8%; Score 8; DB 8; Length 535;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811
 |||||
 Db 55 GLVEGKGP 62

RESULT 91

ADJ50119
 ID ADJ50119 standard; protein; 535 AA.

XX
 AC ADJ50119;

XX
 DT 06-MAY-2004 (first entry)

XX Oil-associated gene related protein #1619.

DE oil-associated gene; transgenic; enhanced seed oil; vegetable oil.

XX

XX Unidentified.

XX US2004025202-A1.

XX 05-FEB-2004.

XX 14-MAR-2003; 2003US-00389566.

XX 15-MAR-2002; 2002US-0365301P.

PR 26-JUN-2002; 2002US-0391786P.

PR 26-JUN-2002; 2002US-0392018P.

XX (LAUR/) LAURIE C C.

PA (RAVA/) RAVANELLO M.

PA (SAVA/) SAVAGE T.

PA (LEDE/) LEDEAUX J R.

PA (ROGE/) ROGERS J A.

XX Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;

XX WPI; 2004-142683/14.

XX Novel recombinant DNA construct comprising a promoter functional in
 PT plants operably linked to an oil-associated gene for producing transgenic
 PT plant seed.

XX Example 3; SEQ ID NO 2123; 22pp; English.

XX The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in plants operably linked to an oil-associated gene.
 CC The construct is useful for transgenic plant seed which has in its genome
 CC the construct, that is functional in the plant to transcribe the oil-
 CC associated gene. The transgenic plant seed grows into a plant having
 CC enhanced seed oil as compared to wild type. The construct is useful for
 CC producing hybrid maize seed. The transgenic plant seed is useful for
 CC producing vegetable oil. The present sequence represents the amino acid
 CC sequence of an oil-associated gene related protein.

SQ Sequence 535 AA;

Query Match 0.8%; Score 8; DB 8; Length 535;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811
 |||||
 Db 55 GLVEGKGP 62

RESULT 92

ADJ49037
 ID ADJ49037 standard; protein; 535 AA.

XX
 AC ADJ49037;

XX
 DT 06-MAY-2004 (first entry)

XX Oil-associated gene related protein #537.

DE oil-associated gene; transgenic; enhanced seed oil; vegetable oil.

XX

XX Unidentified.

XX US2004025202-A1.

XX 05-FEB-2004.

XX 14-MAR-2003; 2003US-00389566.

XX 15-MAR-2002; 2002US-0365301P.

PR 26-JUN-2002; 2002US-0391786P.

PR 26-JUN-2002; 2002US-0392018P.

XX (LAUR/) LAURIE C C.

PA (RAVA/) RAVANELLO M.

PA (SAVA/) SAVAGE T.

PA (LEDE/) LEDEAUX J R.

PA (ROGE/) ROGERS J A.

XX Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;

XX WPI; 2004-142683/14.

XX Novel recombinant DNA construct comprising a promoter functional in
 PT plants operably linked to an oil-associated gene for producing transgenic
 PT plant seed.

XX Example 3; SEQ ID NO 1041; 22pp; English.

XX The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in plants operably linked to an oil-associated gene.
 CC The construct is useful for transgenic plant seed which has in its genome
 CC the construct, that is functional in the plant to transcribe the oil-
 CC associated gene. The transgenic plant seed grows into a plant having
 CC enhanced seed oil as compared to wild type. The construct is useful for
 CC producing hybrid maize seed. The transgenic plant seed is useful for
 CC producing vegetable oil. The present sequence represents the amino acid
 CC sequence of an oil-associated gene related protein.

XX Sequence 535 AA;

Query Match 0.8%; Score 8; DB 8; Length 535;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811
 |||||
 Db 55 GLVEGKGP 62

RESULT 93

ADJ50182
 ID ADJ50182 standard; protein; 535 AA.

XX
 AC ADJ50182;

XX
 DT 06-MAY-2004 (first entry)

XX DE Oil-associated gene related protein #1682.
 XX KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
 XX OS Unidentified.
 XX OS US2004025202-A1.
 XX PD 05-FEB-2004.
 XX XX 14-MAR-2003; 2003US-00389566.
 XX XX 15-MAR-2002; 2002US-0365301P.
 XX PR 26-JUN-2002; 2002US-0391786P.
 XX PR 26-JUN-2002; 2002US-0392018P.
 XX XX (LAUR/) LAURIE C C.
 XX PA (RAVA/) RAVANELLO M.
 XX PA (SAVA/) SAVAGE T.
 XX PA (LEDE/) LEDBAUX J R.
 XX PA (ROGE/) ROGERS J A.
 XX XX Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
 XX PI WPI; 2004-142683/14.
 XX DR Novel recombinant DNA construct comprising a promoter functional in
 XX PT plants operably linked to an oil-associated gene for producing transgenic
 XX PT plant seed.
 XX XX Example 3; SEQ ID NO 2186; 22pp; English.
 XX XX The invention relates to a recombinant DNA construct comprising a
 XX CC promoter functional in plants operably linked to an oil-associated gene.
 XX CC The construct is useful for transgenic plant seed which has in its genome
 XX CC the construct, that is functional in the plant to transcribe the oil-
 XX CC associated gene. The transgenic plant seed grows into a plant having
 XX CC enhanced seed oil as compared to wild type. The construct is useful for
 XX CC producing hybrid maize seed. The transgenic plant seed is useful for
 XX CC producing vegetable oil. The present sequence represents the amino acid
 XX CC sequence of an oil-associated gene related protein.
 XX SQ Sequence 535 AA;
 Query Match 0.8%; Score 8; DB 8; Length 535;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 804 GLVEGKGP 811
 DB 55 GLVEGKGP 62
 RESULT 94
 AAB07724
 ID AAB07724 standard; protein; 555 AA.
 AC AAB07724;
 XX DT 07-NOV-2000 (first entry)
 XX DE An Arabidopsis aintegumenta (ANT) polypeptide.
 XX KW Aintegumenta; ANT; cell proliferation; growth; organ mass; fertility;
 XX KW asexual reproduction; plant; male sterile plant; female sterile plant;
 XX KW early flowering.
 XX OS Arabidopsis sp.
 XX XX Key Location/Qualifiers
 XX FH 281..357
 XX FT Domain /note= "AP2 domain"

FT Domain 383..451
 FT /note= "AP2 domain"
 XX WO200040694-A2.
 XX PD 13-JUL-2000.
 XX PF 07-JAN-2000; 2000WO-US000465.
 XX PR 08-JAN-1999; 99US-00227421.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI Fischer RL, Mizukami Y;
 XX XX WPI; 2000-465969/40.
 XX DR N-PSDB; AA59220.
 XX XX Modulating growth and cell proliferation in a plant used to alter organ
 XX PT mass, control fertility and enhance asexual reproduction in plants
 XX PT comprises modulating ANT activity and selecting plants with altered cell
 XX PT number.
 XX PS Claim 11; Page 41; 54pp; English.
 XX CC The present sequence represents an aintegumenta (ANT) polypeptide. The
 XX CC ANT gene is expressed and functions not only in developing ovules but
 XX CC also in various developing organs. Growth and cell proliferation in plant
 XX CC can be modulated by modulating ANT activity. Modulation of ANT activity
 XX CC is used to alter organ mass, control fertility and enhance asexual
 XX CC reproduction in plants. Increased ANT activity can be used to produce
 XX CC male or female sterile plants. Inhibition of ANT activity can be used to
 XX CC truncate vegetative growth, resulting in early flowering
 XX SQ Sequence 555 AA;
 Query Match 0.8%; Score 8; DB 3; Length 555;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 LKTVALTL 12
 DB 536 LKTVALTL 543
 RESULT 95
 AAG31380
 ID AAG31380 standard; protein; 555 AA.
 XX AC AAG31380;
 XX DT 17-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 37675.
 XX KW Protein identification; signal transduction pathway; metabolic pathway;
 XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
 XX KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX XX EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-00301439.
 XX PR 25-FEB-1999; 99US-0121825P.
 XX PR 05-MAR-1999; 99US-0123180P.
 XX PR 09-MAR-1999; 99US-0123548P.
 XX PR 23-MAR-1999; 99US-0125788P.
 XX PR 25-MAR-1999; 99US-0126264P.
 XX PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131443P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135623P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140693P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142053P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.

PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 0.8%; Score 8; DB 3; Length 555;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12
 |||||
 Db 536 LKTVALTTL 543

RESULT 96
 ABR40832
 ID ABR40832 standard; protein; 555 AA.
 XX
 AC ABR40832;

16-MAY-2003 (first entry)

DE Arabidopsis thaliana oil trait related protein sequence SEQ ID NO:424.

XX Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;
 KW receptor-like protein kinase; mitogen activated protein kinase; oil;
 KW Lip15-like transcription factor caleosin; ATP citrate lyase; SNF1;
 KW CKC-like transcription factor; antisense inhibition; co-suppression;
 KW transgenic plant.

OS Arabidopsis thaliana.

XX WO2003002751-A2.

PN 09-JAN-2003.

PD 27-JUN-2002; 2002WO-US020152.

PF 29-JUN-2001; 2001US-0301913P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

PA (PION-) PIONEER HI-BRED INT INC.

XX Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvell LT;
 PI Jones TJ, Kinney AJ, Klein TM, Li C, Oliveira IC, Sakai H, Shen B;
 PI Tarczynski MC;

XX WPI; 2003-201509/19.

XX

PT Novel nucleotide fragment encoding polypeptides having receptor-like
 PT protein kinase activity, caleosin-like activity, useful for altering oil
 PT phenotypes in plants such as sunflower, coconut, soybean, wheat and rice.
 XX Claim 12; Page 463-465; 542pp; English.

XX The present invention describes an isolated nucleotide fragment (I)
 CC comprising a nucleic acid sequence (NS) chosen from a NS encoding a
 CC polypeptide (PP) having receptor-like protein kinase activity, mitogen
 CC activated protein (MAP)-kinase activity, Lip15-like transcription factor
 CC activity, caleosin-like activity, ATP citrate lyase activity, SNF1-like
 CC activity and CKC-like transcription factor activity. Also described: (1)
 CC complement (III) of (I); (2) a chimeric construct (III) comprising (I) or
 CC (II), operably linked to a regulatory sequence; (3) a plant (IV)
 CC comprising (III) in its genome; (4) seeds (V) obtained from (IV); and (5)
 CC oil obtained from (V). (I) or its part can be used in antisense
 CC inhibition or co-suppression in a transformed plant. (III) is useful for
 CC altering the oil phenotype in a plant such as corn, soybean, wheat, rice,
 CC canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for
 CC creating transgenic plants having altered lipid profiles. (I) can also be
 CC used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to
 CC ABR40879 represent sequences used in the exemplification of the present
 CC invention

XX Sequence 555 AA;

Query Match 0.8%; Score 8; DB 6; Length 555;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12
 |||||
 Db 536 LKTVALTTL 543

RESULT 97

ABR40830

ID ABR40830 standard; protein; 555 AA.

XX

AC ABR40830;

XX 16-MAY-2003 (first entry)

XX Arabidopsis thaliana oil trait related protein sequence SEQ ID NO:422.

XX Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;
 KW receptor-like protein kinase; mitogen activated protein kinase; oil;
 KW Lip15-like transcription factor caleosin; ATP citrate lyase; SNF1;
 KW CKC-like transcription factor; antisense inhibition; co-suppression;
 KW transgenic plant.

OS Arabidopsis thaliana.

XX WO2003002751-A2.

PN 09-JAN-2003.

PD 27-JUN-2002; 2002WO-US020152.

PF 29-JUN-2001; 2001US-0301913P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

PA (PION-) PIONEER HI-BRED INT INC.

XX Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvell LT;
 PI Jones TJ, Kinney AJ, Klein TM, Li C, Oliveira IC, Sakai H, Shen B;
 PI Tarczynski MC;

XX WPI; 2003-201509/19.

PT Novel nucleotide fragment encoding polypeptides having receptor-like
 PT protein kinase activity, caleosin-like activity, useful for altering oil
 PT phenotypes in plants such as sunflower, coconut, soybean, wheat and rice.

XX Claim 12; Page 460-461; 542pp; English.

PS The present invention describes an isolated nucleotide fragment (I) comprising a nucleic acid sequence (NS) chosen from a NS encoding a polypeptide (PP) having receptor-like protein kinase activity, mitogen activated protein (MAP)-kinase activity, LIP15-like transcription factor activity, calceosin-like activity, ATP citrate lyase activity, SNF1-like activity, and CKC-like transcription factor activity. Also described: (1) complement (II) of (I); (2) a chimeric construct (III) comprising (I) or (II), operably linked to a regulatory sequence; (3) a plant (IV) comprising (III) in its genome; (4) seeds (V) obtained from (IV); and (5) oil obtained from (V). (I) or its part can be used in antisense inhibition or co-suppression in a transformed plant. (III) is useful for altering the oil phenotype in a plant such as corn, soybean, wheat, rice, canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for creating transgenic plants having altered lipid profiles. (I) can also be used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to ABR40879 represent sequences used in the exemplification of the present invention

XX Sequence 555 AA;

Query Match 0.8%; Score 8; DB 6; Length 555;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVALTL 12
Db 536 LKTVALTL 543
|||||

RESULT 98
ADG25136
ID ADG25136 standard; protein; 555 AA.

AC ADG25136;

XX 26-FEB-2004 (first entry)

XX Thalecress Aintegumenta, ANT.

XX Thalecress; Aintegumenta; ANT; AP2 domain; plant; seed mass; asexual reproduction; reduced fertility.

XX Arabidopsis thaliana.

XX US6639128-B1

XX 28-OCT-2003.

XX 07-JAN-2000; 2000US-00479855.

XX 08-JAN-1999; 99US-00227421.

XX (NASC-) NAT SCI FOUND.

XX Fischer RL, Mizukami Y;

XX WPI; 2003-842795/78.

XX N-PSDB; ADG25135.

XX New nucleic acid, useful in conferring desired traits on plants, such as increased seed mass, asexual reproduction or reduced fertility.

XX Disclosure; SEQ ID NO 2; 25pp; English.

XX The invention relates to an isolated nucleic acid comprising a polynucleotide encoding a polypeptide having at least 90% identity with Canola Aintegumenta, ANT, (appearing as ADG25139) as determined using a BLAST algorithm and comprising two AP2 domains. Also included are an expression cassette comprising a plant promoter operably linked to a heterologous polynucleotide encoding the polypeptide, a vector comprising

CC the expression cassette, a plant comprising the plant promoter operably linked to the heterologous polynucleotide, a method of asexually reproducing a plant and a method of reducing fertility in a plant. The nucleic acid is useful in conferring desired traits on plants, such as increased seed mass, asexual reproduction or reduced fertility. The present sequence represents Thalecress Aintegumenta, ANT.

XX Sequence 555 AA;

Query Match 0.8%; Score 8; DB 7; Length 555;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVALTL 12
Db 536 LKTVALTL 543
|||||

RESULT 99
ADG88451
ID ADG88451 standard; protein; 555 AA.

AC ADG88451;

XX 11-MAR-2004 (first entry)

XX Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANTMRII.

XX Plant; genetic engineering; cell proliferation; aintegumenta; ANT; transgenic; transgenic plant; organ mass alteration; fertility; asexual reproduction; mutant; muten.

XX Synthetic.

XX Arabidopsis thaliana.

XX US2003159180-A1.

XX 21-AUG-2003.

XX 28-JAN-2002; 2002US-00059911.

XX 28-JAN-2002; 2002US-00059911.

XX (REGC) UNIV CALIFORNIA OFFICE TECHNOLOGY.

XX Fischer RL, Mizukami Y;

XX WPI; 2003-787370/74.

XX N-PSDB; ADG88468.

XX Increasing cell proliferation in a plant by introducing into the plant an expression cassette containing a plant promoter operably linked to a nucleic acid encoding a modified ANT polypeptide and selecting plants with increased mass.

XX Claim 18; SEQ ID NO 22; Sipp; English.

XX The present invention relates to plant genetic engineering. The invention particularly relates to a method of increasing cell proliferation in a plant. The method involves introducing into the plant an expression cassette containing a plant promoter operably linked to a nucleic acid encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2 domain and selecting plants the with increased size or mass. The invention is useful to generate transgenic plants. The method is useful for increasing cell proliferation in a plant for altering organ mass, controlling fertility or enhancing asexual reproduction. The present sequence is Arabidopsis thaliana ANT mutant protein. This sequence is used in the exemplification of the invention.

XX Sequence 555 AA;

Query Match 0.8%; Score 8; DB 7; Length 555;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Search completed: August 28, 2005, 11:04:51
Job time : 177 secs

Matches 8; :Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTL 12
| | | | |
Db 536 LKTVALTL 543

RESULT 100

ADG88431
ID ADG88431 standard; protein; 555 AA.

XX AC ADG88431;

XX DT 11-MAR-2004 (first entry)

XX DE Arabidopsis thaliana aintegumenta (ANT) protein.

XX KW Plant; genetic engineering; cell proliferation; aintegumenta; ANT;
KW transgenic; transgenic plant; organ mass alteration; fertility;
KW asexual reproduction.

XX OS Arabidopsis thaliana.

XX FH Key Location/Qualifiers

FT Domain 281..357

FT FT /note = AP2 domain

FT Region 358..382

FT FT /note = "Linker peptide"

FT Domain 383..451

FT FT /note = AP2 domain

XX PN US2003159180-A1.

XX XX

XX PD

XX PF

XX PP 28-JAN-2002; 2002US-00059911.

XX PR 28-JAN-2002; 2002US-00059911.

XX PA (REGC) UNIV CALIFORNIA OFFICE TECHNOLOGY.

XX PI Fischer RL, Mizukami Y;

XX DR WPI; 2003-787370/74.

XX DR N-PSDB; ADG88430.

XX PT Increasing cell proliferation in a plant by introducing into the plant an
PT expression cassette containing a plant promoter operably linked to a
PT nucleic acid encoding a modified ANT polypeptide and selecting plants
PT with increased mass.

XX PS Example 1; SEQ ID NO 2; 51pp; English.

XX CC The present invention relates to plant genetic engineering. The invention
CC particularly relates to a method of increasing cell proliferation in a
CC plant. The method involves introducing into the plant an expression
CC cassette containing a plant promoter operably linked to a nucleic acid
CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2
CC domain and selecting plants the with increased size or mass. The
CC invention is useful to generate transgenic plants. The method is useful
CC for increasing cell proliferation in a plant for altering organ mass,
CC controlling fertility or enhancing asexual reproduction. The present
CC sequence is Arabidopsis thaliana ANT protein.

XX SQ Sequence 555 AA;

Query Match 0.8%; Score 8; DB 7; Length 555;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTL 12

| | | | |

Db 536 LKTVALTL 543

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 11:08:08 ; Search time 169 Seconds

(without alignments)
2343.607 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 1008

Sequence: 1 MKHLKTVALTITVSVVTH.....LGYTSVALLSLTAIKKKY 1008

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1767149 seqs, 392926209 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1767149

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 110 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1008	100.0	1008	14 US-10-078-531-2	Sequence 2, Appli
2	971	96.3	971	14 US-10-078-531-7	Sequence 7, Appli
3	907	90.0	1008	16 US-10-474-792-46	Sequence 46, Appli
4	667	66.2	969	14 US-10-078-531-8	Sequence 8, Appli
5	610	60.5	970	14 US-10-078-531-4	Sequence 4, Appli
6	540	53.6	971	14 US-10-078-531-6	Sequence 6, Appli
7	465	46.1	963	14 US-10-078-531-5	Sequence 5, Appli
8	263	26.1	951	14 US-10-078-531-3	Sequence 3, Appli
9	76	7.5	1055	14 US-10-091-007-24	Sequence 24, Appli
10	76	7.5	1055	16 US-10-476-614-2	Sequence 2, Appli
11	21	2.1	21	14 US-10-078-531-13	Sequence 13, Appli

279	16	0.9	9	12	US-10-408-765A-935	Sequence 935, Appl
583	15	0.9	9	13	US-10-210-281-34	Sequence 34, Appl
608	13	0.9	9	14	US-10-087-192-549	Sequence 549, Appl
883	15	0.9	9	15	US-10-295-027-484	Sequence 484, Appl
883	15	0.9	9	16	US-10-295-027-484	Sequence 1330, Ap
883	15	0.9	9	17	US-10-173-939-28	Sequence 28, Appl
907	10	0.9	9	18	US-09-965-536A-16	Sequence 16, Appl
907	10	0.9	9	19	US-09-970-944-26	Sequence 26, Appl
907	10	0.9	9	20	US-09-970-944-26	Sequence 27, Appl
907	14	0.9	9	21	US-10-251-385-264	Sequence 264, Appl
907	14	0.9	9	22	US-10-251-385-278	Sequence 278, Appl
907	14	0.9	9	23	US-10-225-567A-422	Sequence 422, Appl
907	15	0.9	9	24	US-10-295-027-849	Sequence 849, Appl
907	15	0.9	9	25	US-10-295-027-946	Sequence 946, Appl
907	15	0.9	9	26	US-10-295-027-1331	Sequence 1331, Ap
907	16	0.9	9	27	US-10-751-736-84	Sequence 84, Appl
907	17	0.9	9	28	US-10-482-029-158	Sequence 158, Appl
1145	17	0.9	9	29	US-10-505-486-103	Sequence 103, Appl
56	15	0.8	8	30	US-10-424-599-178222	Sequence 178222,
82	15	0.8	8	31	US-10-424-599-245882	Sequence 245882,
131	16	0.8	8	32	US-10-425-115-263516	Sequence 263516,
142	16	0.8	8	33	US-10-425-115-306321	Sequence 306321,
179	16	0.8	8	34	US-10-425-115-331445	Sequence 331445,
183	15	0.8	8	35	US-10-424-599-280877	Sequence 280877,
196	14	0.8	8	36	US-10-059-911-12	Sequence 12, Appl
237	17	0.8	8	37	US-10-732-923-5727	Sequence 5727, Ap
239	16	0.8	8	38	US-10-425-115-262882	Sequence 262882,
264	14	0.8	8	39	US-10-059-911-24	Sequence 24, Appl
275	14	0.8	8	40	US-10-059-911-25	Sequence 25, Appl
277	14	0.8	8	41	US-10-059-911-11	Sequence 11, Appl
308	14	0.8	8	42	US-10-059-911-10	Sequence 10, Appl
356	15	0.8	8	43	US-10-282-122A-45121	Sequence 45121, A
411	14	0.8	8	44	US-10-332-585-7284	Sequence 7284, Ap
458	15	0.8	8	45	US-10-369-493-17401	Sequence 17401, A
473	14	0.8	8	46	US-10-059-911-9	Sequence 9, Appli
480	18	0.8	8	47	US-10-481-032A-532	Sequence 532, App
488	15	0.8	8	48	US-10-389-566-712	Sequence 712, App
488	15	0.8	8	49	US-10-389-566-731	Sequence 731, App
488	15	0.8	8	50	US-10-389-566-1366	Sequence 1366, Ap
488	15	0.8	8	51	US-10-389-566-1367	Sequence 1367, Ap
496	18	0.8	8	52	US-10-481-032A-104	Sequence 104, App
500	15	0.8	8	53	US-10-389-566-2286	Sequence 2286, Ap
503	15	0.8	8	54	US-10-389-566-860	Sequence 860, App
503	15	0.8	8	55	US-10-389-566-1411	Sequence 1411, Ap
503	15	0.8	8	56	US-10-389-566-2188	Sequence 2188, Ap
527	15	0.8	8	57	US-10-389-566-859	Sequence 859, App
533	15	0.8	8	58	US-10-389-566-1020	Sequence 1020, Ap
533	15	0.8	8	59	US-10-389-566-1021	Sequence 1021, Ap
535	15	0.8	8	60	US-10-389-566-1041	Sequence 1041, Ap
535	15	0.8	8	61	US-10-389-566-1124	Sequence 1124, Ap
535	15	0.8	8	62	US-10-389-566-2123	Sequence 2123, Ap
535	15	0.8	8	63	US-10-389-566-2186	Sequence 2186, Ap
535	15	0.8	8	64	US-10-389-566-2397	Sequence 2397, Ap
555	14	0.8	8	65	US-10-059-911-2	Sequence 2, Appli
555	14	0.8	8	66	US-10-059-911-21	Sequence 21, Appl
555	14	0.8	8	67	US-10-059-911-22	Sequence 22, Appl
555	15	0.8	8	68	US-10-183-687-422	Sequence 422, App
555	15	0.8	8	69	US-10-183-687-424	Sequence 424, App
555	18	0.8	8	70	US-10-623-477-2	Sequence 2, Appli
642	16	0.8	8	71	US-10-739-930-6762	Sequence 6762, Ap
660	9	0.8	8	72	US-09-864-761-47959	Sequence 47959, A
686	10	0.8	8	73	US-09-981-397A-22	Sequence 22, Appl
686	18	0.8	8	74	US-10-756-149-4945	Sequence 4945, Ap
688	16	0.8	8	75	US-10-638-941-102	Sequence 102, App
715	15	0.8	8	76	US-10-425-114-54382	Sequence 54382, A
820	14	0.8	8	77	US-10-029-386-32324	Sequence 32324, A
879	17	0.8	8	78	US-10-732-923-13561	Sequence 13561, A
916	16	0.8	8	79	US-10-473-574-22	Sequence 22, Appl
931	15	0.8	8	80	US-10-170-385-39	Sequence 39, Appl
931	15	0.8	8	81	US-10-408-765A-1585	Sequence 1585, Ap
976	9	0.8	8	82	US-09-969-528-2	Sequence 2, Appli
3878	14	0.8	8	83	US-10-080-608A-11	Sequence 11, Appl
3899	14	0.8	8	84	US-10-171-311-4	Sequence 4, Appli

85 8 0.8 3907 14 US-10-171-311-2 Sequence 2, Appli
86 8 0.8 3908 18 US-10-756-149-5739 Sequence 5739, Ap
87 8 0.8 3911 15 US-10-370-685-1000 Sequence 100, App
88 8 0.8 3911 16 US-10-408-765A-1839 Sequence 1839, Ap
89 8 0.8 3917 14 US-10-171-311-8 Sequence 8, Appli
90 8 0.8 3925 14 US-10-171-311-6 Sequence 6, Appli
91 7 0.7 16 13 US-10-024-918-8 Sequence 8, Appli
92 7 0.7 16 13 US-10-024-918-12 Sequence 12, Appli
93 7 0.7 16 14 US-10-325-021-13 Sequence 13, Appli
94 7 0.7 16 15 US-10-650-509-8 Sequence 8, Appli
95 7 0.7 16 15 US-10-650-509-12 Sequence 12, Appli
96 7 0.7 21 16 US-10-343-663A-36 Sequence 36, Appli
97 7 0.7 41 9 US-09-864-761-45902 Sequence 45902, A
98 7 0.7 46 15 US-10-424-599-179769 Sequence 179769,
99 7 0.7 47 15 US-10-424-599-185551 Sequence 185551,
100 7 0.7 51 15 US-10-424-599-225766 Sequence 225766,
101 7 0.7 52 16 US-10-425-115-285715 Sequence 285715,
102 7 0.7 52 16 US-10-425-115-340870 Sequence 340870,
103 7 0.7 56 15 US-10-424-599-194696 Sequence 194696,
104 7 0.7 57 16 US-10-425-115-247846 Sequence 247846,
105 7 0.7 58 15 US-10-424-599-231448 Sequence 231448,
106 7 0.7 63 16 US-10-425-115-217832 Sequence 217832,
107 7 0.7 66 16 US-10-425-115-327358 Sequence 327358,
108 7 0.7 70 9 US-09-867-550-364 Sequence 364, App
109 7 0.7 70 16 US-10-425-115-249445 Sequence 249445,
110 7 0.7 71 15 US-10-424-599-211164 Sequence 211164,

ALIGNMENTS

RESULT 1
US-10-078-531-2
; Sequence 2, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE OF INVENTION: DNA FRAGMENTS
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1008
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-2

Query Match 100.0%; Score 1008; DB 14; Length 1008;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKKHLKTVALTITTSVTHNQEVSLVKEPIKQTOASSISGADYAESGSKLKINE 60
DB 1 MKKHLKTVALTITTSVTHNQEVSLVKEPIKQTOASSISGADYAESGSKLKINE 60
QY 61 TSGPVDVDTDLFSKRTTPEKIKNLAKGPREQELKAVTENSEKQITSGSQLEQSK 120
DB 61 TSGPVDVDTDLFSKRTTPEKIKNLAKGPREQELKAVTENSEKQITSGSQLEQSK 120
QY 121 SLSLNKTVPSTNWEICDFTKGNLTVLGSKGVEKLSQTDHLVLPQAADGTQLIQVAS 180
DB 121 SLSLNKTVPSTNWEICDFTKGNLTVLGSKGVEKLSQTDHLVLPQAADGTQLIQVAS 180
QY 181 FAFTPDKKTAAEYTSRAGENGESQLDVGKEIINEGEVFNLSLLKKVTIPTGYKHIGQ 240

DB 181 FAFTPDKKTAAEYTSRAGENGESQLDVGKEIINEGEVFNLSLLKKVTIPTGYKHIGQ 240
QY 241 DAFVNDKNIAEVLNPESLETISDYAFAPHAHLAKQIDLDPNLKAIGELAFQNFQITGKLSLP 300
DB 241 DAFVNDKNIAEVLNPESLETISDYAFAPHAHLAKQIDLDPNLKAIGELAFQNFQITGKLSLP 300
QY 301 RQMLRLAERAPKSNHKTIEFRGNSLKVIGBASQDNDLSQMLPDGKLEKTESAFQNGP 360
DB 301 RQMLRLAERAPKSNHKTIEFRGNSLKVIGBASQDNDLSQMLPDGKLEKTESAFQNGP 360
QY 361 GDDHNNRVVLTWTSKGNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB 361 GDDHNNRVVLTWTSKGNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
QY 421 GFSNKGLOKQVRKNLNLBIPOKHNGVTITEGDNAFRNVDFONKTLRKLYDLEVKLPSTIR 480
DB 421 GFSNKGLOKQVRKNLNLBIPOKHNGVTITEGDNAFRNVDFONKTLRKLYDLEVKLPSTIR 480
QY 481 KIGAPAFQSNLKSFEASDDLEEKEGAFMNNRLETLELKDQVITIGDAAPHINHIYAIV 540
DB 481 KIGAPAFQSNLKSFEASDDLEEKEGAFMNNRLETLELKDQVITIGDAAPHINHIYAIV 540
QY 541 LPESVQSIGRSAPFRQNGANNLI FMGSKVKTILGEMAFLSNRLEHLDLSEKQLTPIVQAF 600
DB 541 LPESVQSIGRSAPFRQNGANNLI FMGSKVKTILGEMAFLSNRLEHLDLSEKQLTPIVQAF 600
QY 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSIAFNALDDNDGDSQFQNKVVVK 660
DB 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSIAFNALDDNDGDSQFQNKVVVK 660
QY 661 THNSYALADGEHFIIVDPDKLSSTIVDLKILKLEGLDYSTLRQTQTQFRDMTTAGKA 720
DB 661 THNSYALADGEHFIIVDPDKLSSTIVDLKILKLEGLDYSTLRQTQTQFRDMTTAGKA 720
QY 721 LLSKSNLRQGEKQFLOEAOFFLGRVLDLKAIAEAKALVTKATKNGQLLERSINKAVL 780
DB 721 LLSKSNLRQGEKQFLOEAOFFLGRVLDLKAIAEAKALVTKATKNGQLLERSINKAVL 780
QY 781 AYNSAIKKNVKKRLEKELDLTLGLVEGKGLAQATVWQVYLLKTPPLPEYIYGLNVY 840
DB 781 AYNSAIKKNVKKRLEKELDLTLGLVEGKGLAQATVWQVYLLKTPPLPEYIYGLNVY 840
QY 841 FDKSGKLIYALDMSDDTTEGQKDAYGNPILNVDNEDNEGYHALAVATLADYEGLDIKTILN 900
DB 841 FDKSGKLIYALDMSDDTTEGQKDAYGNPILNVDNEDNEGYHALAVATLADYEGLDIKTILN 900
QY 901 SKLSQLTSIROVPTAAVHRAGIFQAIQNAAEASQLLPKPGTHSEKSSSSANSKDRGL 960
DB 901 SKLSQLTSIROVPTAAVHRAGIFQAIQNAAEASQLLPKPGTHSEKSSSSANSKDRGL 960
QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYIGLYTSVALLSLITAIKKKKY 1008
DB 961 QSNPKTNRGRHSAILPRTGSKGSFVYIGLYTSVALLSLITAIKKKKY 1008

RESULT 2
US-10-078-531-7
; Sequence 7, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE OF INVENTION: DNA FRAGMENTS
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21


```
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-7

Query Match      96.3%; Score 971; DB 14; Length 971;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LVKEPILKQTOASSISGADYAESGSKLKINETSGPVDVDTDLFSDKRTTPEKIKON 86
DB 1 LVKEPILKQTOASSISGADYAESGSKLKINETSGPVDVDTDLFSDKRTTPEKIKON 60
QY 87 LAKGPREQELKAVTENTSESKQITSGSQLEQESLSLNTKVPSTSNWEICDPITKGNL 146
DB 61 LAKGPREQELKAVTENTSESKQITSGSQLEQESLSLNTKVPSTSNWEICDPITKGNL 120
QY 147 VGLSGSGVEKLSQTDHLVPSQAADGTQLIQVASFAPFTDPKKTAAIAEYTSRAGENGESIQ 206
DB 121 VGLSGSGVEKLSQTDHLVPSQAADGTQLIQVASFAPFTDPKKTAAIAEYTSRAGENGESIQ 180
QY 207 LDVDGKEIINEGEVFNFSYLLKKVTIPTGYKHIGQDAFVQNNKIAEVLNLSIDYAP 266
DB 181 LDVDGKEIINEGEVFNFSYLLKKVTIPTGYKHIGQDAFVQNNKIAEVLNLSIDYAP 240
QY 267 AHLAKQIDLPDLNKAIGELAFDQNTGKLSLPRQLMRLAERAPKSNHIKITEFRGNSL 326
DB 241 AHLAKQIDLPDLNKAIGELAFDQNTGKLSLPRQLMRLAERAPKSNHIKITEFRGNSL 300
QY 327 KVTGEASFQNDLSQLMDPLDGLKEIESEAFNGPDHYNRNVVLWTKSGKNPSGLATEN 386
DB 301 KVTGEASFQNDLSQLMDPLDGLKEIESEAFNGPDHYNRNVVLWTKSGKNPSGLATEN 360
QY 387 TYNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVKENKLEIPKQNGVT 446
DB 361 TYNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVKENKLEIPKQNGVT 420
QY 447 ITEIGNAPRNVDQNKTLKYDLBEVKLPSTTRKIGAFAPQSNLKSFEASDDLEIYE 506
DB 421 ITEIGNAPRNVDQNKTLKYDLBEVKLPSTTRKIGAFAPQSNLKSFEASDDLEIYE 480
QY 507 GAPMNNRIETLEKDKLVITGDAAFINHIYATVLPESVQEIERSAFRONGANNLIFMGS 566
DB 481 GAPMNNRIETLEKDKLVITGDAAFINHIYATVLPESVQEIERSAFRONGANNLIFMGS 540
QY 567 KVKTGEMAFSLNRLHDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKKNH 626
DB 541 KVKTGEMAFSLNRLHDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKKNH 600
QY 627 LKQLEVASALSHITAFNALDNDGDEQFNKVVVTHNSYALADGBHFIVDPDKLSSTIV 686
DB 601 LKQLEVASALSHITAFNALDNDGDEQFNKVVVTHNSYALADGBHFIVDPDKLSSTIV 660
QY 687 DLEKILKLEGLDYSTLROTTQFRDMMTAGKALLSKNLRQGEKQKFLQEAQFELGRV 746
DB 661 DLEKILKLEGLDYSTLROTTQFRDMMTAGKALLSKNLRQGEKQKFLQEAQFELGRV 720
QY 747 DLDKATAKAEKALVTKKATNGQLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLV 806
DB 721 DLDKATAKAEKALVTKKATNGQLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLV 780
QY 807 EGKGPLAQATWQGVYLLKTPPLPEYIYGLNYPYFDKSGKLIYALDMSDTIGEGKDAYG 866
DB 781 EGKGPLAQATWQGVYLLKTPPLPEYIYGLNYPYFDKSGKLIYALDMSDTIGEGKDAYG 840
QY 867 NPILNVDNEGHALAVATLADYEGLDIKTILNSKLSQTSIRQVPTAAYHRAFIQAI 926
DB 841 NPILNVDNEGHALAVATLADYEGLDIKTILNSKLSQTSIRQVPTAAYHRAFIQAI 900
QY 927 QNAAAEAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFYV 986
```

```

901 QNAAAEAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFYV 960
987 GILGYTSVALL 997
961 GILGYTSVALL 971

RESULT 3
US-10-474-792-46
; Sequence 46, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 46
; LENGTH: 1008
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-46

Query Match      90.0%; Score 907; DB 16; Length 1008;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1007; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKKHLKTVALTTLTTVVVTHNQEVFSLVKEPILKQTOASSISGADYAESGSKSKLKINE 60
DB 1 MKKHLKTVALTTLTTVVVTHNQEVFSLVKEPILKQTOASSISGADYAESGSKSKLKINE 60
QY 61 TSGPVDVDTVDLFSKRTTPEKIKONLAGPREQELKAVTENTSESKQITSGSQLEQESKE 120
DB 61 TSGPVDVDTVDLFSKRTTPEKIKONLAGPREQELKAVTENTSESKQITSGSQLEQESKE 120
QY 121 SLSLNTKVPSTSNWEICDPITKGNLTVGLSGSGVEKLSQTDHLVPSQAADGTQLIQVAS 180
DB 121 SLSLNTKVPSTSNWEICDPITKGNLTVGLSGSGVEKLSQTDHLVPSQAADGTQLIQVAS 180
QY 181 FAPTPDKKTAAIAEYTSRAGENGESIQLDVDGKEIINEGEVFNFSYLLKKVTIPTGYKHIGQ 240
DB 181 FAPTPDKKTAAIAEYTSRAGENGESIQLDVDGKEIINEGEVFNFSYLLKKVTIPTGYKHIGQ 240
QY 241 DAFVQNNKIAEVLNLSIDYAFAPHLAKQIDLPDLNKAIGELAFDQNTGKLSL 300
DB 241 DAFVQNNKIAEVLNLSIDYAFAPHLAKQIDLPDLNKAIGELAFDQNTGKLSL 300
QY 301 RQMLRAERAFKSNHIKITEFRGNSLKVIGEASFQNDLSQLMDPLDGLKEIESEAFNGP 360
DB 301 RQMLRAERAFKSNHIKITEFRGNSLKVIGEASFQNDLSQLMDPLDGLKEIESEAFNGP 360
QY 361 GDHYNRNVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB 361 GDHYNRNVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
QY 421 GFSNKGLOKVKENKLEIPKQNGVTITEIGNAFRNVDQNKTLKYDLBEVKLPSTTR 480
DB 421 GFSNKGLOKVKENKLEIPKQNGVTITEIGNAFRNVDQNKTLKYDLBEVKLPSTTR 480
QY 481 KIGAFAPQSNLKSFEASDDLEIEKEGAFMNNRIETLEKDKLVITGDAAFINHIYATV 540
DB 481 KIGAFAPQSNLKSFEASDDLEIEKEGAFMNNRIETLEKDKLVITGDAAFINHIYATV 540
QY 541 LPESVQEIERSAFRONGANNLIFMGSVKKTGEMAFSLNRLHDLSEKQLTEIPVQAF 600
DB 541 LPESVQEIERSAFRONGANNLIFMGSVKKTGEMAFSLNRLHDLSEKQLTEIPVQAF 600
```

QY 601 SDNALKEVLLPASLTIREAAFKKHLKQLEVASALSHTAFNALDDNDGDEQFNKVYK 660
DB 601 SDNALKEVLLPASLTIREAAFKKHLKQLEVASALSHTAFNALDDNDGDEQFNKVYK 660
QY 661 THNSYALADGSHFIVDPDKLSSTIVDLKILKLEGLDYSLRQTQTQFRDMTTAGKA 720
DB 661 THNSYALADGSHFIVDPDKLSSTIVDLKILKLEGLDYSLRQTQTQFRDMTTAGKA 720
QY 721 LLSKNLROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGQLLERSINKAVL 780
DB 721 LLSKNLROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGQLLERSINKAVL 780
QY 781 AYNSAIAKKANYKLEKELDLTLGLVEGKPLAQATMVGVLKATPLPEYIIGLVNY 840
DB 781 AYNSAIAKKANYKLEKELDLTLGLVEGKPLAQATMVGVLKATPLPEYIIGLVNY 840
QY 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDEGYPHALAVATLADYEGLDIKTILN 900
DB 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDEGYPHALAVATLADYEGLDIKTILN 900
QY 901 SKLSQTSIRQVPTAAHYRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGL 960
DB 901 SKLSQTSIRQVPTAAHYRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGL 960
QY 961 QSNPKTNRGRHSAILPRTGSKGSPVYIGLGYTSVALLSLITAIKKKKY 1008
DB 961 QSNPKTNRGRHSAILPRTGSKGSPVYIGLGYTSVALLSLITAIKKKKY 1008

RESULT 4

US-10-078-531-8
; Sequence 8, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-8

Query Match 66.2%; Score 667; DB 14; Length 969;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 867; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 128 VPSTSNWEICDITKGNLTIVGLSGVEKLSQTDHLVLPQAAADGTQLIQVAFPAFTDK 187
DB 101 VPSTSNWEICDITKGNLTIVGLSGVEKLSQTDHLVLPQAAADGTQLIQVAFPAFTDK 160
QY 198 KTAIAEYTSRAGENGESOLDVDGKEIINEGVFNSYLLKKVTIPTGYKHIGQDAFVNK 247
DB 161 KTAIAEYTSRAGENGESOLDVDGKEIINEGVFNSYLLKKVTIPTGYKHIGQDAFVNK 220
QY 248 NTAENVLPESLTIIDYAPAHALAKQIDLDPNLKAIGELAFPDNQITGKLSIPROLMELA 307
DB 221 NTAENVLPESLTIIDYAPAHALAKQIDLDPNLKAIGELAFPDNQITGKLSIPROLMELA 280
QY 308 BRAFKSNHIKTIETFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNPGDDHNN 367
DB 281 BRAFKSNHIKTIETFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNPGDDHNN 340

QY 368 RVVLWTKSGKPNPSGLATENTYVNPDKSLWQSSPIDYTKMLEEDFTYQKNSVTGFSNKG 427
DB 341 RVVLWTKSGKPNPSGLATENTYVNPDKSLWQSSPIDYTKMLEEDFTYQKNSVTGFSNKG 400
QY 428 QKVRKNKLEIPKOHNGVTITEIGDAPRNVDFONKTLRKYDLEEVKLPSITIRKIGAFAP 487
DB 401 QKVRKNKLEIPKOHNGVTITEIGDAPRNVDFONKTLRKYDLEEVKLPSITIRKIGAFAP 460
QY 488 QSNNLKSPFASDDLEEKEGAPMNNRIETLEBKDLVTIGDAAAPHINHIYAIVLPEVQE 547
DB 461 QSNNLKSPFASDDLEEKEGAPMNNRIETLEBKDLVTIGDAAAPHINHIYAIVLPEVQE 520
QY 548 IGRSAFRONGANNLI FMGSKVKTIGEMAFISNRLEHLDLSEKQKLTETIPVQAFSDNALKE 607
DB 521 IGRSAFRONGANNLI FMGSKVKTIGEMAFISNRLEHLDLSEKQKLTETIPVQAFSDNALKE 580
QY 608 VLLPASLKTIREBAFKKHLKQLEVASALSHTAFNALDDNDGDEQFNKVYVVKTHNSYA 667
DB 591 VLLPASLKTIREBAFKKHLKQLEVASALSHTAFNALDDNDGDEQFNKVYVVKTHNSYA 640
QY 668 LADGEHFIVDPDKLSSTIVDLKILKLEGLDYSLRQTQTQFRDMTTAGKALLSKSNL 727
DB 641 LADGEHFIVDPDKLSSTIVDLKILKLEGLDYSLRQTQTQFRDMTTAGKALLSKSNL 700
QY 728 RQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAI 787
DB 701 RQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAI 760
QY 788 KKANVKRLEKELDLTLGLVEGKPLAQATMVGVLKATPLPEYIIGLVNYFDKSGKL 847
DB 761 KKANVKRLEKELDLTLGLVEGKPLAQATMVGVLKATPLPEYIIGLVNYFDKSGKL 820
QY 848 IYALDMSDTIGEGQKDAYGNPILNVDEDEGYPHALAVATLADYEGLDIKTILSKLSQ 907
DB 821 IYALDMSDTIGEGQKDAYGNPILNVDEDEGYPHALAVATLADYEGLDIKTILSKLSQ 880
QY 908 SIROVPTAAHYRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGLQSNPKTN 967
DB 881 SIROVPTAAHYRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKDRGLQSNPKTN 940
QY 968 RGRHSAILPRTGSKGSPVYIGLGYTSVAL 996
DB 941 RGRHSAILPRTGSKGSPVYIGLGYTSVAL 969

RESULT 5

US-10-078-531-4
; Sequence 4, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 970
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-4

Query Match 60.5%; Score 610; DB 14; Length 970;
Best Local Similarity 99.8%; Pred. No. 0;


```
Db          913 KPG 915

RESULT 7
US-10-078-531-5
; Sequence 5, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 963
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-5

Query Match      46.1%; Score 465; DB 14; Length 963;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 865; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 128 VPSTSNWEICDFTITKGNLTVLGSLGSGVEKLSQTDHLVLPQQAADGTQLIQVAFPTDPK 187
Db          913 KPG 915
QY 95 VPSTSNWEICDFTITKGNLTVLGSLGSGVEKLSQTDHLVLPQQAADGTQLIQVAFPTDPK 154
Db          913 KPG 915
QY 188 KTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTGKHHIGQDAFVDNK 247
Db          913 KPG 915
QY 155 KTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTGKHHIGQDAFVDNK 214
QY 248 NIAEVLNLPESLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSLPRQLMRLA 307
Db          913 KPG 915
QY 215 NIAEVLNLPESLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSLPRQLMRLA 274
QY 308 ERAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFPTGNPGDDHNN 367
Db          913 KPG 915
QY 275 ERAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFPTGNPGDDHNN 334
QY 368 RVVLTWTKSGKNPSGLATENTYNNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKG 427
Db          913 KPG 915
QY 335 RVVLTWTKSGKNPYGLATENTYNNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKG 394
QY 428 QKVAKNKLPIKQHGVTITEIGDAPFNVDVFNKTLRKDYDLBEVKLPSTIRKIGAFAP 487
Db          913 KPG 915
QY 395 QKVAKNKLPIKQHGVTITEIGDAPFNVDVFNKTLRKDYDLBEVKLPSTIRKIGAFAP 454
QY 488 QSNNLKSPFASDDEBEIKEGAFMNNRIETLEKDKLVITIGDAFPHNHIYAIVLPSVQ 547
Db          913 KPG 915
QY 455 QSNNLKSPFASDDEBEIKEGAFMNNRIETLEKDKLVITIGDAFPHNHIYAIVLPSVQ 514
QY 548 IGRSAPFRONGANNLIIPMGSKVTKLGEMAFSLNRLHDLSEKQLETEIPVQAFSNALKE 607
Db          913 KPG 915
QY 515 IGRSAPFRONGANNLIIPMGSKVTKLGEMAFSLNRLHDLSEKQLETEIPVQAFSNALKE 574
QY 608 VLLPASLKTIREAFKKNHLKQLEVASALSHTAFNALDNDGDEQFNKVVVKTTHNSYA 667
Db          913 KPG 915
QY 575 VLLPASLKTIREAFKKNHLKQLEVASALSHTAFNALDNDGDEQFNKVVVKTTHNSYA 634
QY 668 LADGSHFIVDPDKLSSTIVDLBKILKLEGLDYSLRTQTQTPRDMTTAGKALLSKNL 727
Db          913 KPG 915
QY 635 LADGSHFIVDPDKLSSTIVDLBKILKLEGLDYSLRTQTQTPRDMTTAGKALLSKNL 694
QY 728 ROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTTKATNGQLLERSINKAVIAYNSAI 787

Db          913 KPG 915
QY 695 ROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTTKATNGQLLERSINKAVIAYNSAI 754
QY 788 KKANVKRLKELDLITGLVEGKPLAQATWVQGVYLLKTLPLPLPEYITGLNVYFDKSGKL 847
Db          913 KPG 915
QY 755 KKANVKRLKELDLITGLVEGKPLAQATWVQGVYLLKTLPLPLPEYITGLNVYFDKSGKL 814
QY 848 IYALDMSDTIGEGQKDAYGNPILNVDSBNBYHALAVATLADYEGLODKITLNSKLSQUT 907
Db          913 KPG 915
QY 815 IYALDMSDTIGEGQKDAYGNPILNVDSBNBYHALAVATLADYEGLODKITLNSKLSQUT 874
QY 908 SIROVPTAAVHRAGIFQAIQNAAAEQQLPKPTHSEKSSSSSANSKDRGLQSNPKTN 967
Db          913 KPG 915
QY 875 SIROVPTAAVHRAGIFQAIQNAAAEQQLPKPTHSEKSSSSSANSKDRGLQSNPKTN 934
QY 968 RGRHSAILPRTGSGKSFVYGLGYTSVAL 996
Db          913 KPG 915
QY 935 RGRHSAILPRTGSGKSFVYGLGYTSVAL 963

RESULT 8
US-10-078-531-3
; Sequence 3, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 951
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-3

Query Match      26.1%; Score 263; DB 14; Length 951;
Best Local Similarity 99.3%; Pred. No. 1.8e-243;
Matches 863; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 128 VPSTSNWEICDFTITKGNLTVLGSLGSGVEKLSQTDHLVLPQQAADGTQLIQVAFPTDPK 187
Db          913 KPG 915
QY 83 VPSTSNWEICDFTITKGNLTVLGSLGSGVEKLSQTDHLVLPQQAADGTQLIQVAFPTDPK 142
QY 188 KTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTGKHHIGQDAFVDNK 247
Db          913 KPG 915
QY 143 KTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTGKHHIGQDAFVDNK 202
QY 248 NIAEVLNLPESLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSLPRQLMRLA 307
Db          913 KPG 915
QY 203 NIAEVLNLPESLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSLPRQLMRLA 262
QY 308 ERAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFPTGNPGDDHNN 367
Db          913 KPG 915
QY 263 ERAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFPTGNPGDDHNN 322
QY 368 RVVLTWTKSGKNPSGLATENTYNNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKG 427
Db          913 KPG 915
QY 323 RVVLTWTKSGKNPYGLATENTYNNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKG 382
QY 428 QKVAKNKLPIKQHGVTITEIGDAPFNVDVFNKTLRKDYDLBEVKLPSTIRKIGAFAP 487
Db          913 KPG 915
QY 383 QKVAKNKLPIKQHGVTITEIGDAPFNVDVFNKTLRKDYDLBEVKLPSTIRKIGAFAP 442
```

Qy	488	QSNLKSFEASDDLEBIEKEGAFMNNRIETLELKKDLVTIGDAAFHNIHIAIVLPESVOE	547
Db	443	QSNLKSFEASDDLEBIEKEGAFMNNRIETLELKKDLVTIGDAAFHNIHIAIVLPESVOE	502
Qy	548	IGRSAFRONGANNLIFMGSKVKITLGEMAFLSNRLSHLDLSEKQUTEIPVQAFPSNALKE	607
Db	503	IGRSAFRONGANNLIFMGSKVKITLGEMAFLSNRLSHLDLSEKQUTEIPVQAFPSNALKE	562
Qy	608	VLLPASLKTIREEAFKQHLKQLEVASALSHLAFNALDDNDGDEQFDNKVVVKTHNSYA	667
Db	563	VLLPASLKTIREEAFKQHLKQLEVASALSHLAFNALDDNDGDEQFDNKVVVKTHNSYA	622
Qy	668	LADGEHFIVDPDKLSSITVLDLEKILKLEGLDYSTLRQTQTQFRDMTTAGKALLSKSL	727
Db	623	LADGEHFIVDPDKLSSITVLDLEKILKLEGLDYSTLRQTQTQFRDMTTAGKALLSKSL	682
Qy	728	ROGEKQKPLQBAQPFPLGRVLDLDKATAKAELVTKKATKNQGLLERSINKAVLAYNNSAI	787
Db	683	ROGEKQKPLQBAQPFPLGRVLDLDKATAKAELVTKKATKNQGLLERSINKAVLAYNNSAI	742
Qy	788	KKANVKRLEKELDLTLTGVLGEGKPLQAQTMVQGVYLLKTPPLPEYITGLNVYFDKSGKL	847
Db	743	KKANVKRLEKELDLTLTGVLGEGKPLQAQTMVQGVYLLKTPPLPEYITGLNVYFDKSGKL	802
Qy	848	IYALDMSDTIGEGQKDAYGNPFLNVDENEGYPHALAVATLADYEGLDIKTILNSKLSOLT	907
Db	803	IYALDMSDTIGEGQKDAYGNPFLNVDENEGYPHALAVATLADYEGLDIKTILNSKLSOLT	862
Qy	908	SIRQVPTAAHYHRAGIFQAIQNAAAEAEQLPKPGTHSEKSSSESANGKDRGLQSNPKTN	967
Db	863	SIRQVPTAAHYHRAGIFQAIQNAAAEAEQLPKPGTHSEKSSSESANGKDRGLQSNPKTN	922
Qy	968	RGHSAILPRTGSGKGSFVYIGLYTYSVAL	996
Db	923	RGHSAILPRTGSGKGSFVYIGLYTYSVAL	951

```

RESULT 9
US-10-091-007-24
; Sequence 24, Application US/10091007
; Publication No. US20030170782A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics limited
; APPLICANT: Le Page, Richard W F Wells, Jeremy M
; APPLICANT: Hannifry, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21978WO
; CURRENT APPLICATION NUMBER: US/10/091,007
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: GB 9921125.2
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 24
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-091-007-24

```

```

Query Match          7.5%; Score 76; DB 14; Length 1055;
Best Local Similarity 100.0%; Pred. No. 2e-63;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy      805  LVGKGPLAQTWQGVYLLKTPPLPBYIYGLNVPYFDKSGKLIYALDMSDTIGSQKDA 864
        |||
Db      806  LVGKGPLAQTWQGVYLLKTPPLPBYIYGLNVPYFDKSGKLIYALDMSDTIGSQKDA 865

Qy      865  YGNPILNVDEDNQYH 880
        |||
Db      866  YGNPILNVDEDNQYH 881
        |||

RESULT 10

```

```

US-10-476-614-2
; Sequence 2, Application US/10476614
; Publication No. US2004017113A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem Inc.
; TITLE OF INVENTION: Antigenes of Group B Streptococcus and Corresponding DNA Fragment-
; FILE REFERENCE: 74872-81
; CURRENT APPLICATION NUMBER: US/10/476,614
; CURRENT FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: US 60/287,712
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Streptococcus
US-10-476-614-2

```

Query Match	7.5%	Score 76;	DB 16;	Length 1055;
Best Local Similarity	100.0%;	Pred. No. 28-63;		
Matches	76;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0
Qy	805	LVEKGGLPQAQATWVGQVYLLKTPLPPEYYITGLNVYFDKSGKLIYALDMSDTIGSQKDA	864	
Db	806	LVEKGGLPQAQATWVGQVYLLKTPLPPEYYITGLNVYFDKSGKLIYALDMSDTIGSQKDA	865	
Qy	865	YGNPILNVDEDNEGYH	880	
Db	866	YGNPILNVDEDNEGYH	881	

```

RESULT 11
US-10-078-531-13
; Sequence 13, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEF
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Signal peptide
US-10-078-531-13

```

```
Query Match      2.1%; Score 21; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.5e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```



```
QY      1 MKKHLKTVALTLTTSVVVTHN 21
          |||||
Db       1 MKKHLKTVALTLTTSVVVTHN 21
```



```
RESULT 12
US-10-408-765A-935
; Sequence 935, Application US/10408765A
; Publication No. US20040101874A1
GENERAL INFORMATION:
```

; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Faby, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
 ; FILE REFERENCE: 66088.465
 ; CURRENT APPLICATION NUMBER: US/10/408,765A
 ; CURRENT FILING DATE: 2003-04-04
 ; NUMBER OF SEQ ID NOS: 3077
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 935
 ; LENGTH: 279
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-408-765A-935

Query Match 0.9%; Score 9; DB 16; Length 279;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952
 Db 18 SEKSSSES 26

RESULT 13
 US-10-210-281-34
 ; Sequence 34, Application US/10210281
 ; Publication No. US20040030096A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gorman, Linda
 ; APPLICANT: Zethusen, Bryan D.
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Guo, Xiaojia
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: Zhong, Mei
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Miller, Charles E.
 ; APPLICANT: Jil Weizhen
 ; APPLICANT: Pena, Carol E.A.
 ; APPLICANT: Bugesse, Catherine E.
 ; APPLICANT: Sciore, Paul
 ; APPLICANT: Stone, David J.
 ; APPLICANT: Taupier, Raymond J., Jr.
 ; APPLICANT: Cahan, Stacie
 ; APPLICANT: Rothenberg, Mark E.
 ; APPLICANT: Malyankar, Uriel M.
 ; APPLICANT: Boldog, Ferenc L.
 ; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
 ; TITLE OF INVENTION: THE SAME
 ; FILE REFERENCE: 21402-416D
 ; CURRENT APPLICATION NUMBER: US/10/210,281
 ; CURRENT FILING DATE: 2003-02-05
 ; PRIOR APPLICATION NUMBER: 60/309,501
 ; PRIOR FILING DATE: 2001-08-02
 ; PRIOR APPLICATION NUMBER: 60/310,291
 ; PRIOR FILING DATE: 2001-08-03
 ; PRIOR APPLICATION NUMBER: 60/361,775
 ; PRIOR FILING DATE: 2002-03-05
 ; PRIOR APPLICATION NUMBER: 60/310,951
 ; PRIOR FILING DATE: 2001-08-08
 ; PRIOR APPLICATION NUMBER: 60/361,832
 ; PRIOR FILING DATE: 2002-03-05
 ; PRIOR APPLICATION NUMBER: 60/311,292
 ; PRIOR FILING DATE: 2001-08-09
 ; PRIOR APPLICATION NUMBER: 60/311,979
 ; PRIOR FILING DATE: 2001-08-13
 ; PRIOR APPLICATION NUMBER: 60/312,203

; PRIOR FILING DATE: 2001-08-14
 ; PRIOR APPLICATION NUMBER: 60/313,201
 ; PRIOR FILING DATE: 2001-08-17
 ; PRIOR APPLICATION NUMBER: 60/313,702
 ; PRIOR FILING DATE: 2001-08-20
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 191
 ; SOFTWARE: CuraseqList version 0.1
 ; SEQ ID NO 34
 ; LENGTH: 583
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-210-281-34

Query Match 0.9%; Score 9; DB 15; Length 583;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952
 Db 18 SEKSSSES 26

RESULT 14
 US-10-087-192-549
 ; Sequence 549, Application US/10087192
 ; Publication No. US20020182586A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, David W.
 ; APPLICANT: Engelhard, Eric K.
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
 ; TITLE OF INVENTION: CANCER
 ; FILE REFERENCE: 529452000122
 ; CURRENT APPLICATION NUMBER: US/10/087,192
 ; CURRENT FILING DATE: 2002-03-01
 ; PRIOR APPLICATION NUMBER: US 09/747,377
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: US 09/798,586
 ; PRIOR FILING DATE: 2001-03-02
 ; NUMBER OF SEQ ID NOS: 2059
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 549
 ; LENGTH: 608
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-087-192-549

Query Match 0.9%; Score 9; DB 13; Length 608;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952
 Db 36 SEKSSSES 44

RESULT 15
 US-10-295-027-484
 ; Sequence 484, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glynn, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Eos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

```

; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 484
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-484

```

```

Query Match      0.9%; Score 9; DB 15; Length 883;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      592 LTEIPVQAF 600
Db      174 LTEIPVQAF 182

```

```

RESULT 16
US-10-295-027-1330
; Sequence 1330, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14

```

```

; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1330
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1330

```

```

Query Match      0.9%; Score 9; DB 15; Length 883;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      592 LTEIPVQAF 600
Db      174 LTEIPVQAF 182

```

```

RESULT 17
US-10-173-999-28
; Sequence 28, Application US/10173999
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002420US
; CURRENT APPLICATION NUMBER: US/10/173,999
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/299,234
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-173-999-28

```

```

Query Match      0.9%; Score 9; DB 15; Length 883;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      592 LTEIPVQAF 600
Db      174 LTEIPVQAF 182

```

```

RESULT 18
US-09-965-536A-16
; Sequence 16, Application US/09965536A
; Publication No. US20030027323A1
; GENERAL INFORMATION:
; APPLICANT: FEDER, J. N.
; APPLICANT: MINTIER, G.
; APPLICANT: RAMANATHAN, C. S.

```

; APPLICANT: HAWKEN, D. R.
; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMV5,
; FILE REFERENCE: D0041NP
; CURRENT APPLICATION NUMBER: US/09/965,536A
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,713
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/261,781
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/306,605
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/310,436
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 907
; TYPE: PRT
; ORGANISM: HUMAN
US-09-965-536A-16

Query Match 0.9%; Score 9; DB 10; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 19
US-09-970-944-26
; Sequence 26, Application US/09970944
; Publication No. US20030204052A1
; GENERAL INFORMATION:
; APPLICANT: Herrman, John L
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: No. US20030204052A1el Proteins and Nucleic Acids Encoding Same an
; TITLE OF INVENTION: Antibodies Directed Against these Proteins
; FILE REFERENCE: 21402-138
; CURRENT APPLICATION NUMBER: US/09/970,944
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: 60/237,862
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-970-944-26

Query Match 0.9%; Score 9; DB 10; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 20
US-09-970-944-27
; Sequence 27, Application US/09970944
; Publication No. US20030204052A1
; GENERAL INFORMATION:
; APPLICANT: Herrman, John L
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: No. US20030204052A1el Proteins and Nucleic Acids Encoding Same an
; TITLE OF INVENTION: Antibodies Directed Against these Proteins

; FILE REFERENCE: 21402-138
; CURRENT APPLICATION NUMBER: US/09/970,944
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: 60/237,862
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-970-944-27

Query Match 0.9%; Score 9; DB 10; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 21
US-10-251-385-264
; Sequence 264, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 264
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-251-385-264

Query Match 0.9%; Score 9; DB 14; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 22
US-10-251-385-278
; Sequence 278, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 278
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-251-385-278

Query Match 0.9%; Score 9; DB 14; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
||| |||||
Db 174 LTEIPVQAF 182

RESULT 23

US-10-225-567A-422
; Sequence 422, Application US/10225567A
; Publication No. US2003013798A1

; GENERAL INFORMATION:

; APPLICANT: LifeSpan Biosciences

; APPLICANT: Brown, Joseph P.

; APPLICANT: Burner, Glenna C.

; APPLICANT: Rouse, Christine L.

; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS

; FILE REFERENCE: 1920-4-4

; CURRENT APPLICATION NUMBER: US/10/225,567A

; CURRENT FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/257,144

; PRIOR FILING DATE: 2000-12-19

; NUMBER OF SEQ ID NOS: 2292

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 422

; LENGTH: 907

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-225-567A-422

Query Match 0.9%; Score 9; DB 14; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
||| |||||
Db 174 LTEIPVQAF 182

RESULT 24

US-10-295-027-849
; Sequence 849, Application US/10295027
; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glynn, Richard

; APPLICANT: Hevezi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 849
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-849

Query Match 0.9%; Score 9; DB 15; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600
||| |||||
Db 174 LTEIPVQAF 182

RESULT 25

US-10-295-027-946
; Sequence 946, Application US/10295027
; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glynn, Richard

; APPLICANT: Hevezi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464

; PRIOR FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US 60/334,393

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: US 60/340,376

; PRIOR FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: US 60/347,211

; PRIOR FILING DATE: 2002-01-08

; PRIOR APPLICATION NUMBER: US 60/347,349

; PRIOR FILING DATE: 2002-01-10

; PRIOR APPLICATION NUMBER: US 60/355,250

; PRIOR FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: US 60/356,714

; PRIOR FILING DATE: 2002-02-13

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 946
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-946

Query Match 0.9%; Score 9; DB 15; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 26

US-10-295-027-1331
; Sequence 1331, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1331
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1331

Query Match 0.9%; Score 9; DB 15; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 27

US-10-751-736-84
; Sequence 84, Application US/10751736
; Publication No. US20040265230A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Martinez, Robert
; APPLICANT: Brown, Eugene
; APPLICANT: Liu, Wei
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING AND TREATING COLON
; FILE REFERENCE: AM100927 (031896-002000)
; CURRENT APPLICATION NUMBER: US/10/751,736
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US Provisional Application 60/438,000
; PRIOR FILING DATE: 2003-01-06
; NUMBER OF SEQ ID NOS: 54873
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 84
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-751-736-84

Query Match 0.9%; Score 9; DB 16; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 28

US-10-482-029-158
; Sequence 158, Application US/10482029
; Publication No. US20050037445A1
; GENERAL INFORMATION:
; APPLICANT: ODIN medical A/S
; TITLE OF INVENTION: Oncology drug innovation
; FILE REFERENCE: P 573 PC00
; CURRENT APPLICATION NUMBER: US/10/482,029
; CURRENT FILING DATE: 2003-12-29
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 158
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-482-029-158

Query Match 0.9%; Score 9; DB 17; Length 907;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

RESULT 29

US-10-505-486-103
; Sequence 103, Application US/10505486
; Publication No. US20050118639A1
; GENERAL INFORMATION:
; APPLICANT: Takeda Chemical Industries, Ltd.
; TITLE OF INVENTION: Determination of a ligand
; FILE REFERENCE: P03-0006PCT
; CURRENT APPLICATION NUMBER: US/10/505,486
; CURRENT FILING DATE: 2004-08-20
; PRIOR APPLICATION NUMBER: JP 2002-45728

; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: JP 2002-213949
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: JP 2002-298237
; PRIOR FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 233
; SEQ ID NO 103
; LENGTH: 1145
; TYPE: PRT
; ORGANISM: Human
US-10-505-486-103

Query Match 0.9%; Score 9; DB 17; Length 1145;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVOAF 600
Db 174 LTEIPVOAF 182

RESULT 30

US-10-424-599-178222
; Sequence 178222, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 178222
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_131950C.1.pap
US-10-424-599-178222

Query Match 0.8%; Score 8; DB 15; Length 56;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 645 DDNDGDEQ 652
Db 4 DDNDGDEQ 11

RESULT 31

US-10-424-599-245882
; Sequence 245882, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 245882
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_131950C.1.pap

; NAME/KEY: unsure
; LOCATION: (1)..(82)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_64062C.1.pap
US-10-424-599-245882

Query Match 0.8%; Score 8; DB 15; Length 82;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 TGKLSLPR 301
Db 8 TGKLSLPR 15

RESULT 32

US-10-425-115-263516
; Sequence 263516, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 263516
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_17193C.1.pap
US-10-425-115-263516

Query Match 0.8%; Score 8; DB 16; Length 131;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSE 951
Db 5 SEKSSSSE 12

RESULT 33

US-10-425-115-306321
; Sequence 306321, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 306321
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(142)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_42429C.1.pap

US-10-425-115-306321

Query Match 0.8%; Score 8; DB 16; Length 142;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 EKIKDNLA 88
| | | | | | | |
Db 99 EKIKDNLA 106

RESULT 34

US-10-425-115-331445
; Sequence 331445, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 331445
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_65381C.1.pgp
US-10-425-115-331445

Query Match 0.8%; Score 8; DB 16; Length 179;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 NETSQPVD 66
| | | | | | | |
Db 139 NETSQPVD 146

RESULT 35

US-10-424-599-280877
; Sequence 280877, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 280877
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(183)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_95654C.1.pgp
US-10-424-599-280877

Query Match 0.8%; Score 8; DB 15; Length 183;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 564 MGSKVKTLL 571
| | | | | | | |
Db 1 MGSKVKTLL 8

RESULT 36

US-10-059-911-12
; Sequence 12, Application US/10059911
; Publication No. US20030159180A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
; FILE REFERENCE: 023070-090730US
; CURRENT APPLICATION NUMBER: US/10/059,911
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:structurally
; OTHER INFORMATION: altered ANT protein ANTIDNA
US-10-059-911-12

Query Match 0.8%; Score 8; DB 14; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVALTLL 12
| | | | | | | |
Db 177 LKTVALTLL 184

RESULT 37

US-10-732-923-5727
; Sequence 5727, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 5727
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Nicotiana tabacum
US-10-732-923-5727

Query Match 0.8%; Score 8; DB 17; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 948 SSESSEANS 955
| | | | | | | |
Db 42 SSESSEANS 49

RESULT 38

US-10-425-115-262882
; Sequence 262882, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 262882
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_17135C.1.pcp
US-10-425-115-262882

Query Match 0.8%; Score 8; DB 16; Length 239;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 EASDDLEE 503
|||||
Db 1 EASDDLEE 8

RESULT 39
US-10-059-911-24
; Sequence 24, Application US/10059911
; Publication No. US20030159180A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
; FILE REFERENCE: 023070-090730US
; CURRENT APPLICATION NUMBER: US/10/059,911
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: novel chimeric
; OTHER INFORMATION: ANT protein ANT-AP25w2
US-10-059-911-24

Query Match 0.8%; Score 8; DB 14; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12
|||||
Db 245 LKTVALTTL 252

RESULT 40
US-10-059-911-25
; Sequence 25, Application US/10059911
; Publication No. US20030159180A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
; FILE REFERENCE: 023070-090730US
; CURRENT APPLICATION NUMBER: US/10/059,911
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 275

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: novel chimeric
; OTHER INFORMATION: ANT protein ANTdf1
US-10-059-911-25

Query Match 0.8%; Score 8; DB 14; Length 275;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12
|||||
Db 256 LKTVALTTL 263

RESULT 41
US-10-059-911-11
; Sequence 11, Application US/10059911
; Publication No. US20030159180A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
; FILE REFERENCE: 023070-090730US
; CURRENT APPLICATION NUMBER: US/10/059,911
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: structurally
; OTHER INFORMATION: altered ANT protein ANTNDN3
US-10-059-911-11

Query Match 0.8%; Score 8; DB 14; Length 277;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12
|||||
Db 258 LKTVALTTL 265

RESULT 42
US-10-059-911-10
; Sequence 10, Application US/10059911
; Publication No. US20030159180A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
; FILE REFERENCE: 023070-090730US
; CURRENT APPLICATION NUMBER: US/10/059,911
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: structurally
; OTHER INFORMATION: altered ANT protein ANTNDN2
US-10-059-911-10

Query Match 0.8%; Score 8; DB 14; Length 308;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVLTL 12
| | | | |
Db 289 LKTVLTL 296

RESULT 43

US-10-282-122A-45121
; Sequence 45121, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45121
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-10-282-122A-45121

Query Match 0.8%; Score 8; DB 15; Length 356;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 997 LSLITAIK 1004
| | | | |
Db 156 LSLITAIK 163

RESULT 44

US-10-032-585-7284
; Sequence 7284, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang

; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7284
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7284

Query Match 0.8%; Score 8; DB 14; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 LEKELDLL 802
| | | | |
Db 37 LEKELDLL 44

RESULT 45

US-10-369-493-17401
; Sequence 17401, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17401
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-10-369-493-17401

Query Match 0.8%; Score 8; DB 15; Length 458;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 KAIKAEK 757
| | | | |
Db 447 KAIKAEK 454

RESULT 46

US-10-059-911-9
; Sequence 9, Application US/10059911
; Publication No. US20030159180A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
; FILE REFERENCE: 023070-090730US
; CURRENT APPLICATION NUMBER: US/10/059,911
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 473
; TYPE: PRT

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:structurally
; OTHER INFORMATION: altered ANT protein ANT0N1
US-10-059-911-9

Query Match          0.8%; Score 8; DB 14; Length 473;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 LKTVALTL 12
        |||||
Db       454 LKTVALTL 461

RESULT 47
US-10-481-032A-532
; Sequence 532, Application US/10481032A
; Publication No. US20050177901A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Cheng, Wengqiong
; APPLICANT: Briggs, Steven
; APPLICANT: Cooper, Bret
; APPLICANT: Goff, Stephen A.
; APPLICANT: Moughamer, Todd
; APPLICANT: Glazebrook, Jane
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Krepes, Joel
; APPLICANT: Provart, Nicolas
; APPLICANT: Rickes, Darrell
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 60148USPCT
; CURRENT APPLICATION NUMBER: US/10/481,032A
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: US 60/300,112
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/342,327
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: PCT/IB02/02450
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 1201
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 532
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-481-032A-532

Query Match          0.8%; Score 8; DB 18; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
        |||||
Db       55 GLVEGKGP 62

RESULT 48
US-10-389-566-712
; Sequence 712, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
```

```
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 712
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-389-566-712

Query Match          0.8%; Score 8; DB 15; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
        |||||
Db       55 GLVEGKGP 62

RESULT 49
US-10-389-566-731
; Sequence 731, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 731
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-389-566-731

Query Match          0.8%; Score 8; DB 15; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
        |||||
Db       55 GLVEGKGP 62

RESULT 50
US-10-389-566-1366
; Sequence 1366, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
```

```
; SOFTWARE: PatehtIn version 3.2
; SEQ ID NO 1366
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-389-566-1366

Query Match          0.8%; Score 8; DB 15; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 51
US-10-389-566-1367
; Sequence 1367, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1367
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-389-566-1367

Query Match          0.8%; Score 8; DB 15; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 52
US-10-481-032A-104
; Sequence 104, Application US/10481032A
; Publication No. US20050177901A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Cheng, Wengdong
; APPLICANT: Briggs, Steven
; APPLICANT: Cooper, Bret
; APPLICANT: Goff, Stephen A.
; APPLICANT: Moughamer, Todd
; APPLICANT: Glazebrook, Jane
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicolas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 60148USPCT
; CURRENT APPLICATION NUMBER: US/10/481,032A
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: US 60/300,112
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/325,277
```

```
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/342,327
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: PCT/IB02/02450
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 1201
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 104
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Oryza sativa
; NAME/KEY: misc feature
; LOCATION: (286)..(286)
; OTHER INFORMATION: Xaa = any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (300)..(300)
; OTHER INFORMATION: Xaa = any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (451)..(451)
; OTHER INFORMATION: Xaa = any naturally occurring amino acid
US-10-481-032A-104

Query Match          0.8%; Score 8; DB 18; Length 496;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 53
US-10-389-566-2286
; Sequence 2286, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2286
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-389-566-2286

Query Match          0.8%; Score 8; DB 15; Length 500;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy      804 GLVEGKGP 811
Db      51 GLVEGKGP 58

RESULT 54
US-10-389-566-860
; Sequence 860, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
```



```

; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 860
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-10-389-566-860

```

```

Query Match          0.8%; Score 8; DB 15; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

```

```

RESULT 55
US-10-389-566-1411
; Sequence 1411, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1411
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-10-389-566-1411

```

```

Query Match          0.8%; Score 8; DB 15; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

```

```

RESULT 56
US-10-389-566-2188
; Sequence 2188, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566

```

```

; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2188
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Secale cereale
US-10-389-566-2188

```

```

Query Match          0.8%; Score 8; DB 15; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

```

```

RESULT 57
US-10-389-566-859
; Sequence 859, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 859
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (19)..(19)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (496)..(497)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-389-566-859

```

```

Query Match          0.8%; Score 8; DB 15; Length 527;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      804 GLVEGKGP 811
Db      57 GLVEGKGP 64

```

```

RESULT 58
US-10-389-566-1020
; Sequence 1020, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C

```

```
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1020
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-389-566-1020

Query Match          0.8%; Score 8; DB 15; Length 533;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      53 GLVEGKGP 60

RESULT 59
US-10-389-566-1021
; Sequence 1021, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1021
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-389-566-1021

Query Match          0.8%; Score 8; DB 15; Length 533;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      53 GLVEGKGP 60

RESULT 60
US-10-389-566-1041
; Sequence 1041, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
```

```
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1041
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-389-566-1041

Query Match          0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 61
US-10-389-566-1124
; Sequence 1124, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1124
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-389-566-1124

Query Match          0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 62
US-10-389-566-2123
; Sequence 2123, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
```

```
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2123
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-389-566-2123

Query Match      0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 63
US-10-389-566-2186
; Sequence 2186, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2186
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-389-566-2186

Query Match      0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 64
US-10-389-566-2397
; Sequence 2397, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2397
; LENGTH: 535
; TYPE: PRT

; ORGANISM: Hordeum vulgare
US-10-389-566-2397

Query Match      0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 65
US-10-059-911-2
; Sequence 2, Application US/10059911
; Publication No. US20030159180A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
; FILE REFERENCE: 023070-090730US
; CURRENT APPLICATION NUMBER: US/10/059,911
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Arabidopsis AINTEGUMENTA (ANT)
US-10-059-911-2

Query Match      0.8%; Score 8; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 LKTVALTL 12
Db      536 LKTVALTL 543

RESULT 66
US-10-059-911-21
; Sequence 21, Application US/10059911
; Publication No. US20030159180A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
; FILE REFERENCE: 023070-090730US
; CURRENT APPLICATION NUMBER: US/10/059,911
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: structurally
; OTHER INFORMATION: altered ANT protein ANTMNLS
US-10-059-911-21

Query Match      0.8%; Score 8; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 LKTVALTL 12
Db      536 LKTVALTL 543
```

```
Best Local Similarity 100.0%; Pred. No. 3.1e+02; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

Qy 5 LKTVLTL 12
Db 536 LKTVLTL 543

RESULT 69
US-10-183-687-424
; Sequence 424, Application US/10183687
; Publication No. US20030204870A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Cahoone, Rebecca
; APPLICANT: Epelbaum, Sabine
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Harvell, Leslie T.
; APPLICANT: Jones, Todd
; APPLICANT: Kinney, Tony
; APPLICANT: Klein, Ted
; APPLICANT: Li, Changjiang
; APPLICANT: Oliveira, Igor Cunha
; APPLICANT: Sakai, Hajime
; APPLICANT: Shen, Bo
; APPLICANT: Tarczynski, Mitchell C.
; TITLE OF INVENTION: Alteration Of Oil Traits In Plants
; FILE REFERENCE: BB1458 US NA
; CURRENT APPLICATION NUMBER: US/10/183,687
; CURRENT FILING DATE: 2002-06-27
; PRIOR APPLICATION NUMBER: 60/301,913
; PRIOR FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 424
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: gi 1171429
US-10-183-687-424

Query Match 0.8%; Score 8; DB 15; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVLTL 12
Db 536 LKTVLTL 543

RESULT 70
US-10-623-477-2
; Sequence 2, Application US/10623477
; Publication No. US2005013245A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass, Controlling Fertility
; FILE REFERENCE: 023070-090720US
; CURRENT APPLICATION NUMBER: US/10/623,477
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US/09/479,855
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 555
```

```
Query Match 0.8%; Score 8; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVLTL 12
Db 536 LKTVLTL 543

RESULT 68
US-10-183-687-422
; Sequence 422, Application US/10183687
; Publication No. US20030204870A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Cahoone, Rebecca
; APPLICANT: Epelbaum, Sabine
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Harvell, Leslie T.
; APPLICANT: Jones, Todd
; APPLICANT: Kinney, Tony
; APPLICANT: Klein, Ted
; APPLICANT: Li, Changjiang
; APPLICANT: Oliveira, Igor Cunha
; APPLICANT: Sakai, Hajime
; APPLICANT: Shen, Bo
; APPLICANT: Tarczynski, Mitchell C.
; TITLE OF INVENTION: Alteration Of Oil Traits In Plants
; FILE REFERENCE: BB1458 US NA
; CURRENT APPLICATION NUMBER: US/10/183,687
; CURRENT FILING DATE: 2002-06-27
; PRIOR APPLICATION NUMBER: 60/301,913
; PRIOR FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 422
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: gi 2129537
US-10-183-687-422

Query Match 0.8%; Score 8; DB 15; Length 555;
```

```
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: AINTEGUMENTA (ANT)
US-10-623-477-2

Query Match      0.8%; Score 8; DB 18; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVLTL 12
Db 536 LKTVLTL 543

RESULT 71
US-10-739-930-6762
; Sequence 6762, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; NUMBER OF SEQ ID NOS: 12-18
; SEQ ID NO 6762
; LENGTH: 642
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C8623_1.p
US-10-739-930-6762
```

```
Query Match      0.8%; Score 8; DB 16; Length 642;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 12 LTVSVVT 19
Db 149 LTVSVVT 156
```

```
RESULT 72
US-09-864-761-47959
; Sequence 47959, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47959
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AJ010770.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q99323, EVALUE 3.00e-17
; OTHER INFORMATION: EST_HUMAN HIT: AU132932.1, EVALUE 1.00e-105
US-09-864-761-47959
```

```
Query Match      0.8%; Score 8; DB 9; Length 660;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 516 TLELKDKL 523
Db 35 TLELKDKL 42
```

```
RESULT 73
US-09-981-397A-22
; Sequence 22, Application US/09981397A
; Publication No. US20030082519A1
; GENERAL INFORMATION:
; APPLICANT: Axxima Pharmaceuticals AG
; APPLICANT: Schubart, Daniel
; APPLICANT: Habenberger, Peter
; APPLICANT: Stein-Gerlach, Matthias
; APPLICANT: Bevec, Dorian
; TITLE OF INVENTION: Cellular Kinases Involved in Cytomegalovirus Infection and their
; FILE REFERENCE: AXM-004.1 US
; CURRENT APPLICATION NUMBER: US/09/981,397A
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: 60/240,750
; PRIOR FILING DATE: 2000-10-16
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 686
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-981-397A-22

Query Match      0.8%; Score 8; DB 10; Length 686;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 944 SEKSSSSE 951
|||||
Db 5 SEKSSSSE 12

RESULT 74

US-10-756-149-4945
; Sequence 4945, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4945
; LENGTH: 686
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-756-149-4945

Query Match 0.8%; Score 8; DB 18; Length 686;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSSE 951
|||||
Db 5 SEKSSSSE 12

RESULT 75

US-10-618-941-102
; Sequence 102, Application US/10618941
; Publication No. US2004019792A1
; GENERAL INFORMATION:
; APPLICANT: WHYTE, DAVID
; APPLICANT: MANNING, GERARD
; APPLICANT: CAENEPEEL, SEAN
; TITLE OF INVENTION: NOVEL KINASES
; FILE REFERENCE: 034536-0321
; CURRENT APPLICATION NUMBER: US/10/618,941
; CURRENT FILING DATE: 2003-07-15
; PRIOR APPLICATION NUMBER: 60/395,632
; PRIOR FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 102
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-618-941-102

Query Match 0.8%; Score 8; DB 16; Length 688;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSSE 951
|||||
Db 5 SEKSSSSE 12

RESULT 76

US-10-425-114-54382
; Sequence 54382, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 54382
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3063-134-E5_FLI.pep
US-10-425-114-54382

Query Match 0.8%; Score 8; DB 15; Length 715;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSSE 951
|||||
Db 32 SEKSSSSE 39

RESULT 77

US-10-029-386-32324
; Sequence 32324, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32324
; LENGTH: 820
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC000065.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: Q99996, EVALUAE 0.00e+00
US-10-029-386-32324

Query Match 0.8%; Score 8; DB 14; Length 820;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKDKL 523
|||||
Db 51 TLELKDKL 58

RESULT 78

US-10-732-923-13561
; Sequence 13561, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C

```
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 13561
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Sycon raphanus
US-10-732-923-13561

Query Match      0.8%; Score 8; DB 17; Length 879;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 KOKFLOEA 739
Db 659 KOKFLOEA 666

RESULT 79
US-10-473-574-22
; Sequence 22, Application US/10473574
; Publication No. US20040116670A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; HAPALIA, April J.A.;
; APPLICANT: TANG, Y. Tom; YUE, Henry;
; APPLICANT: KHAN, Farrah A.; ISON, Craig H.;
; APPLICANT: BAUGHN, Mariah R.; WARREN, Bridget A.;
; APPLICANT: DUGGAN, Brendan M.; THANGAVELU, Kavitha;
; APPLICANT: HONCHELL, Cynthia D.; AZIMZAI, Yalda;
; APPLICANT: ELLIOTT, Vicki S.; BURFORD, Neil;
; APPLICANT: DING, Li; YUE, Huibin;
; APPLICANT: BECHA, Shanya; EMERLING, Brooke M.;
; APPLICANT: RICHARDSON, Thomas W.; LEE, Soo Yeun;
; APPLICANT: BANDMAN, Olga; LAL, Preeti G.;
; APPLICANT: LEE, Sally; GIETZEN, Kimberly J.;
; APPLICANT: CHAWLA, Narinder K.; GRIFFIN, Jennifer A.;
; APPLICANT: LEE, Ernestine A.; SWARNAKAR, Anita;
; APPLICANT: RING, Huijun Z.; JONES, Karen Anne
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0918 USN
; CURRENT APPLICATION NUMBER: US/10/473,574
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US02/09288
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US 60/294,451
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 60/291,870
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/290,518
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/288,609
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/283,769
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/281,323
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/280,508
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 916
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte ID No: 7468507CD1
US-10-473-574-22

Query Match      0.8%; Score 8; DB 16; Length 916;
Best Local Similarity 100.0%; Pred. No. 5e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 KRLKELD 800
Db 707 KRLKELD 714

RESULT 80
US-10-170-385-39
; Sequence 39, Application US/10170385
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 532682000100
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 931
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-170-385-39

Query Match      0.8%; Score 8; DB 15; Length 931;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 945 EKSSSES 952
Db 892 EKSSSES 899

RESULT 81
US-10-408-765A-1585
; Sequence 1585, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Roin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1585
; LENGTH: 931
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1585

Query Match      0.8%; Score 8; DB 16; Length 931;
```

```
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 945 EKSSSSS 952
DB 892 EKSSSSS 899
|||||

RESULT 82
US-09-969-528-2
; Sequence 2, Application US/09969528
; Patent No. US20020150567A1
; GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; Jefferson, Anne Bennett
; Williams, Lewis T.
; Majerus, Philip W.
; TITLE OF INVENTION: NO. US20020150567A1e1 Grb2 Associating Protein and Nucleic
; Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/969,528
; APPLICATION NUMBER: US/09/969,528
; FILING DATE: 01-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/560,005
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-969-528-2
Query Match 0.8%; Score 8; DB 9; Length 976;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 SKLSQLTS 908
DB 50 SKLSQLTS 57
|||||

RESULT 83
US-10-080-608A-11
; Sequence 11, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES

; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-608A-11
Query Match 0.8%; Score 8; DB 14; Length 3878;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKDKL 523
DB 357 TLELKDKL 364
|||||

RESULT 84
US-10-171-311-4
; Sequence 4, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MEI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4
Query Match 0.8%; Score 8; DB 14; Length 3899;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKDKL 523
DB 345 TLELKDKL 352
|||||

RESULT 85
US-10-171-311-2
; Sequence 2, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
```



```
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-2

Query Match      0.8%; Score 8; DB 14; Length 3907;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      516 TLELKDKL 523
      |||||
Db      345 TLELKDKL 352

RESULT 86
US-10-756-149-5739
; Sequence 5739, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5739
; LENGTH: 3908
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-756-149-5739

Query Match      0.8%; Score 8; DB 18; Length 3908;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      516 TLELKDKL 523
      |||||
Db      345 TLELKDKL 352

RESULT 87
US-10-370-685-100
; Sequence 100, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
```

```
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 100
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: human
US-10-370-685-100

Query Match      0.8%; Score 8; DB 15; Length 3911;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      516 TLELKDKL 523
      |||||
Db      357 TLELKDKL 364

RESULT 88
US-10-408-765A-1839
; Sequence 1839, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1839
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1839

Query Match      0.8%; Score 8; DB 16; Length 3911;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      516 TLELKDKL 523
      |||||
Db      357 TLELKDKL 364

RESULT 89
US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
```

```
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-8
```

```
Query Match          0.8%; Score 8; DB 14; Length 3917;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 516 TLELKDKL 523
Db 345 TLELKDKL 352
```

RESULT 90

```
US-10-171-311-6
; Sequence 6, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Horesh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6
```

```
Query Match          0.8%; Score 8; DB 14; Length 3925;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 516 TLELKDKL 523
Db 345 TLELKDKL 352
```

RESULT 91

```
US-10-024-918-8
; Sequence 8, Application US/10024918
; Publication No. US20020168718A1
; GENERAL INFORMATION:
; APPLICANT: Hubbell, Jeffrey
; APPLICANT: Schense, Jason
; APPLICANT: Zisch, Andreas
; APPLICANT: Hall, Heike
; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING
; FILE REFERENCE: CIT 2606 CIP
```

```
; CURRENT APPLICATION NUMBER: US/10/024,918
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 16
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: peptide with glutamine at the transglutaminase coupling site and
; OTHER INFORMATION: an active peptide that mimics the crosslinking site in the gamma
; NAME/KEY: MOD.RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: dansylated tyrosine
US-10-024-918-8
```

```
Query Match          0.7%; Score 7; DB 13; Length 16;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 855 DTIGEGQ 861
Db 4 DTIGEGQ 10
```

RESULT 92

```
US-10-024-918-12
; Sequence 12, Application US/10024918
; Publication No. US20020168718A1
; GENERAL INFORMATION:
; APPLICANT: Hubbell, Jeffrey
; APPLICANT: Schense, Jason
; APPLICANT: Zisch, Andreas
; APPLICANT: Hall, Heike
; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING
; FILE REFERENCE: CIT 2606 CIP
; CURRENT APPLICATION NUMBER: US/10/024,918
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 16
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: peptide with glutamine at the transglutaminase coupling site in t
; OTHER INFORMATION: he chain of fibrinogen
US-10-024-918-12
```

```
Query Match          0.7%; Score 7; DB 13; Length 16;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 855 DTIGEGQ 861
Db 4 DTIGEGQ 10
```

RESULT 93

```
US-10-325-021-13
; Sequence 13, Application US/10325021
; Publication No. US20030166833A1
; GENERAL INFORMATION:
; APPLICANT: Itolf, Matthias
; APPLICANT: Schense, Jason
; APPLICANT: Hubbell, Jeffery Alan
; APPLICANT: Jen, Anna
; TITLE OF INVENTION: Growth Factor Modified Protein Matrices for Tissue
; FILE REFERENCE: ETH 107 CIP(3)
; CURRENT APPLICATION NUMBER: US/10/325,021
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 10/024,918
```

; PRIOR FILING DATE: 2001-12-18
 ; PRIOR APPLICATION NUMBER: PCT/EP02/12458
 ; PRIOR FILING DATE: 2002-11-07
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 13
 ; LENGTH: 16
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-325-021-13

Query Match 0.7%; Score 7; DB 14; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 855 DTIGEGQ 861
 |||||
 Db 4 DTIGEGQ 10

RESULT 94
 US-10-650-509-8
 ; Sequence 8, Application US/10650509
 ; Publication No. US20040082513A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hubbell, Jeffrey
 ; APPLICANT: Schense, Jason
 ; APPLICANT: Zisch, Andreas
 ; APPLICANT: Hall, Heike
 ; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING
 ; FILE REFERENCE: CIT 2606 CIP CON
 ; CURRENT APPLICATION NUMBER: US/10/650,509
 ; CURRENT FILING DATE: 2003-08-27
 ; PRIOR APPLICATION NUMBER: 10/024,918
 ; PRIOR FILING DATE: 2001-12-18
 ; PRIOR APPLICATION NUMBER: 09/057,052
 ; PRIOR FILING DATE: 1998-04-08
 ; PRIOR APPLICATION NUMBER: PCT/US98/06617
 ; PRIOR FILING DATE: 1998-04-02
 ; PRIOR APPLICATION NUMBER: 60/042,143
 ; PRIOR FILING DATE: 1997-04-03
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 8
 ; LENGTH: 16
 ; TYPE: PRT
 ; ORGANISM: artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: peptide with glutamine at the transglutaminase coupling site and
 ; OTHER INFORMATION: an active peptide that mimics the crosslinking site in the gamma
 ; OTHER INFORMATION: chain of fibrinogen
 ; FEATURE:
 ; NAME/KEY: MOD RES
 ; LOCATION: (1)..(1)
 ; OTHER INFORMATION: dansylated tyrosine
 US-10-650-509-8

Query Match 0.7%; Score 7; DB 15; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 855 DTIGEGQ 861
 |||||
 Db 4 DTIGEGQ 10

RESULT 95
 US-10-650-509-12
 ; Sequence 12, Application US/10650509
 ; Publication No. US20040082513A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hubbell, Jeffrey
 ; APPLICANT: Schense, Jason

; APPLICANT: Zisch, Andreas
 ; APPLICANT: Hall, Heike
 ; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING
 ; FILE REFERENCE: CIT 2606 CIP CON
 ; CURRENT APPLICATION NUMBER: US/10/650,509
 ; CURRENT FILING DATE: 2003-08-27
 ; PRIOR APPLICATION NUMBER: 10/024,918
 ; PRIOR FILING DATE: 2001-12-18
 ; PRIOR APPLICATION NUMBER: 09/057,052
 ; PRIOR FILING DATE: 1998-04-08
 ; PRIOR APPLICATION NUMBER: PCT/US98/06617
 ; PRIOR FILING DATE: 1998-04-02
 ; PRIOR APPLICATION NUMBER: 60/042,143
 ; PRIOR FILING DATE: 1997-04-03
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 12
 ; LENGTH: 16
 ; TYPE: PRT
 ; ORGANISM: artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: peptide with glutamine at the transglutaminase coupling site in
 ; OTHER INFORMATION: he chain of fibrinogen
 US-10-650-509-12

Query Match 0.7%; Score 7; DB 15; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 855 DTIGEGQ 861
 |||||
 Db 4 DTIGEGQ 10

RESULT 96
 US-10-343-663A-36
 ; Sequence 36, Application US/10343663A
 ; Publication No. US20040106148A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dowler, S.
 ; APPLICANT: Campbell, D.
 ; APPLICANT: Gray A.
 ; APPLICANT: Downes, P.
 ; APPLICANT: Alessi, D.
 ; TITLE OF INVENTION: Polypeptides
 ; FILE REFERENCE: 002.00220 (MEDY/P24958PC)
 ; CURRENT APPLICATION NUMBER: US/10/343,663A
 ; CURRENT FILING DATE: 2003-12-03
 ; PRIOR APPLICATION NUMBER: PCT/GB01/03481
 ; PRIOR FILING DATE: 2001-08-03
 ; PRIOR APPLICATION NUMBER: GB 0018908.4
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 68
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 36
 ; LENGTH: 21
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Peptide
 US-10-343-663A-36

Query Match 0.7%; Score 7; DB 16; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 194 YTSRAGE 200
 |||||
 Db 2 YTSRAGE 8

RESULT 97
 US-09-864-761-45902

; Sequence 45902, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45902
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008521.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.71
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.77
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.66
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73
US-09-864-761-45902

Query Match 0.7%; Score 7; DB 9; Length 41;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 945 EKSSSSE 951
Db 32 EKSSSSE 38
|||||

RESULT 98

US-10-424-599-179769
; Sequence 179769, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 179769
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_133347C.1.pep
US-10-424-599-179769

Query Match 0.7%; Score 7; DB 15; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 779 VLAYNNS 785
Db 4 VLAYNNS 10
|||||

RESULT 99

US-10-424-599-185551
; Sequence 185551, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 185551
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138567C.1.pep
US-10-424-599-185551

Query Match 0.7%; Score 7; DB 15; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 718 GKALLSK 724
Db 22 GKALLSK 28
|||||

RESULT 100

US-10-424-599-225766
; Sequence 225766, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ.ID NOS: 285684
; SEQ ID NO 225766
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_45899C.1.pap
US-10-424-599-225766

```

```

Query Match      0.7%; Score 7; DB 15; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      427 LQKVKRN 433
      |||||
Db      19 LQKVKRN 25

```

Search completed: August 28, 2005, 11:22:10
Job time : 174 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 11:02:48 ; Search time 43 Seconds
(without alignments)
1749.913 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 1008

Sequence: 1 MKKHLKLTVALTLTVSVVTH.....LGYTSVALLSLTAIKKKY 1008

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 513545 seqs, 74649064 residues

Word size : 0

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 110 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*

2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*

3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*

4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*

5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep:*

6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	0.9	557	4	US-09-568-816A-2
2	9	0.9	907	4	US-09-170-496D-264
3	9	0.9	907	4	US-09-170-496D-278
4	8	0.8	350	2	US-08-852-481-2
5	8	0.8	364	4	US-09-328-352-4367
6	8	0.8	373	4	US-09-248-796A-19382
7	8	0.8	531	1	US-08-531-601-1
8	8	0.8	531	2	US-08-859-032-1
9	8	0.8	535	1	US-08-737-597-10
10	8	0.8	555	4	US-09-227-421-2
11	8	0.8	555	4	US-09-479-855-2
12	8	0.8	688	2	US-09-016-000-2
13	8	0.8	976	3	US-08-560-003-2
14	8	0.8	976	3	US-09-195-868-14
15	8	0.8	976	3	US-09-418-540-2
16	8	0.8	976	4	US-09-969-528-2
17	8	0.8	1187	3	US-08-664-962B-8
18	8	0.8	1187	3	US-09-311-743-8
19	8	0.8	1189	3	US-09-195-868-15
20	8	0.8	1229	3	US-09-195-868-28
21	8	0.8	3878	4	US-09-914-259-11
22	7	0.7	16	3	US-09-057-052-8
23	7	0.7	16	3	US-09-057-052-12
24	7	0.7	16	4	US-09-695-466-8
25	7	0.7	16	4	US-09-695-466-12
26	7	0.7	26	4	US-09-270-767-60016
27	7	0.7	38	4	US-09-270-767-44565

28	7	0.7	39	4	US-09-270-767-61674	Sequence 61674, A
29	7	0.7	60	4	US-09-621-976-6627	Sequence 6627, Ap
30	7	0.7	60	4	US-09-621-976-6631	Sequence 6631, Ap
31	7	0.7	62	4	US-09-248-796A-20152	Sequence 20152, A
32	7	0.7	77	4	US-09-513-999C-5429	Sequence 5429, Ap
33	7	0.7	81	4	US-09-621-976-6626	Sequence 6626, Ap
34	7	0.7	81	4	US-09-621-976-6628	Sequence 6628, Ap
35	7	0.7	81	4	US-09-621-976-6629	Sequence 6629, Ap
36	7	0.7	81	4	US-09-621-976-6630	Sequence 6630, Ap
37	7	0.7	85	4	US-09-248-796A-19130	Sequence 19130, A
38	7	0.7	87	4	US-09-248-796A-16501	Sequence 16501, A
39	7	0.7	90	4	US-09-270-767-40137	Sequence 40137, A
40	7	0.7	90	4	US-09-270-767-55353	Sequence 55353, A
41	7	0.7	94	4	US-09-270-767-44382	Sequence 44382, A
42	7	0.7	101	4	US-09-902-540-11181	Sequence 11181, A
43	7	0.7	105	4	US-09-513-999C-5997	Sequence 5997, Ap
44	7	0.7	106	4	US-09-270-767-44049	Sequence 44049, A
45	7	0.7	111	4	US-09-248-796A-20841	Sequence 20841, A
46	7	0.7	120	4	US-09-071-035-206	Sequence 206, App
47	7	0.7	120	4	US-09-071-035-210	Sequence 210, App
48	7	0.7	121	4	US-09-252-991A-28777	Sequence 28777, A
49	7	0.7	123	4	US-09-270-767-33193	Sequence 33193, A
50	7	0.7	123	4	US-09-270-767-48410	Sequence 48410, A
51	7	0.7	136	4	US-09-270-767-57861	Sequence 57861, A
52	7	0.7	143	4	US-09-134-000C-6044	Sequence 6044, Ap
53	7	0.7	144	3	US-08-858-207A-523	Sequence 523, App
54	7	0.7	147	4	US-09-134-000C-6221	Sequence 6221, Ap
55	7	0.7	152	3	US-08-679-493A-203	Sequence 203, App
56	7	0.7	161	4	US-09-252-991A-24022	Sequence 24022, A
57	7	0.7	172	4	US-09-538-092-450	Sequence 450, App
58	7	0.7	176	4	US-09-270-767-38641	Sequence 38641, A
59	7	0.7	176	4	US-09-270-767-53858	Sequence 53858, A
60	7	0.7	176	4	US-09-893-737-22	Sequence 22, Appl
61	7	0.7	180	4	US-09-270-767-35060	Sequence 35060, A
62	7	0.7	180	4	US-09-270-767-50277	Sequence 50277, A
63	7	0.7	181	4	US-09-270-767-61846	Sequence 61846, A
64	7	0.7	190	4	US-09-902-540-15188	Sequence 15188, A
65	7	0.7	208	4	US-09-270-767-46119	Sequence 46119, A
66	7	0.7	211	4	US-09-248-796A-15823	Sequence 15823, A
67	7	0.7	215	3	US-09-305-984-18	Sequence 18, Appl
68	7	0.7	215	3	US-09-305-984-64	Sequence 64, Appl
69	7	0.7	215	4	US-09-073-541A-18	Sequence 18, Appl
70	7	0.7	215	4	US-09-493-940-18	Sequence 18, Appl
71	7	0.7	215	4	US-09-493-940-64	Sequence 64, Appl
72	7	0.7	215	4	US-09-583-110-5320	Sequence 5320, Ap
73	7	0.7	219	4	US-09-107-433-4810	Sequence 4810, Ap
74	7	0.7	224	3	US-09-305-984-72	Sequence 72, Appl
75	7	0.7	224	3	US-09-305-984-74	Sequence 74, Appl
76	7	0.7	224	3	US-09-305-984-76	Sequence 76, Appl
77	7	0.7	224	4	US-09-493-940-72	Sequence 72, Appl
78	7	0.7	224	4	US-09-493-940-74	Sequence 74, Appl
79	7	0.7	224	4	US-09-493-940-76	Sequence 76, Appl
80	7	0.7	236	4	US-09-902-540-11067	Sequence 11067, A
81	7	0.7	239	4	US-09-198-452A-399	Sequence 399, App
82	7	0.7	239	4	US-09-438-185A-384	Sequence 384, App
83	7	0.7	241	2	US-08-867-087B-11	Sequence 11, Appl
84	7	0.7	241	4	US-09-107-532A-5876	Sequence 5876, Ap
85	7	0.7	242	4	US-09-252-991A-21887	Sequence 21887, A
86	7	0.7	243	4	US-09-328-352-4519	Sequence 4519, Ap
87	7	0.7	248	4	US-09-252-991A-32999	Sequence 32999, A
88	7	0.7	250	4	US-09-270-767-40859	Sequence 40859, A
89	7	0.7	250	4	US-09-270-767-56075	Sequence 56075, A
90	7	0.7	256	4	US-09-248-796A-17359	Sequence 17359, A
91	7	0.7	259	4	US-09-489-039A-7271	Sequence 7271, Ap
92	7	0.7	261	4	US-09-489-039A-11968	Sequence 11968, A
93	7	0.7	268	4	US-09-270-767-42555	Sequence 42555, A
94	7	0.7	270	4	US-09-248-796A-19421	Sequence 19421, A
95	7	0.7	272	4	US-09-489-039A-10410	Sequence 10410, A
96	7	0.7	289	2	US-08-741-437-1	Sequence 1, Appli
97	7	0.7	289	2	US-09-134-593-1	Sequence 1, Appli
98	7	0.7	289	4	US-09-248-796A-18420	Sequence 18420, A
99	7	0.7	289	4	US-09-538-092-1335	Sequence 1335, Ap
100	7	0.7	292	4	US-09-535-521-2	Sequence 2, Appli

101 7 0.7 292 4 US-09-535-521-5 Sequence 5, Appli
102 7 0.7 294 4 US-09-252-991A-26953 Sequence 26953, A
103 7 0.7 295 4 US-09-328-352-4378 Sequence 4378, Ap
104 7 0.7 302 4 US-09-252-991A-21875 Sequence 21875, A
105 7 0.7 306 4 US-09-489-039A-10023 Sequence 10023, A
106 7 0.7 307 4 US-09-489-039A-10116 Sequence 10116, A
107 7 0.7 313 3 US-08-985-335-8 Sequence 8, Appli
108 7 0.7 313 3 US-09-410-372-8 Sequence 8, Appli
109 7 0.7 314 2 US-08-822-701-10 Sequence 10, Appli
110 7 0.7 314 3 US-08-935-855-10 Sequence 10, Appli

ALIGNMENTS

RESULT 1
US-09-568-816A-2
; Sequence 2, Application US/09568816A
; Patent No. 6440599
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Swedlund, Brad
; APPLICANT: Simard, Jacques
; APPLICANT: Rommens, Johanna M.
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Hospital for Sick Children
; TITLE OF INVENTION: CA7 CG04 Gene
; FILE REFERENCE: 2318-237-II
; CURRENT APPLICATION NUMBER: US/09/568,816A
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/134,209
; PRIOR FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-568-816A-2

Query Match 0.9%; Score 9; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSS 952
Db 18 SEKSSSSS 26
RESULT 2
US-09-170-496D-264
; Sequence 264, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 264
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-264

Query Match 0.9%; Score 9; DB 4; Length 907;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVOAF 600
Db 174 LTEIPVOAF 182

RESULT 3

US-09-170-496D-278
; Sequence 278, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 278
; LENGTH: 907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-278

Query Match 0.9%; Score 9; DB 4; Length 907;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVOAF 600
Db 174 LTEIPVOAF 182

RESULT 4

US-08-852-481-2
; Sequence 2, Application US/08852481
; Patent No. 5928931
; GENERAL INFORMATION:
; APPLICANT: Grun Ph.D., Felix
; APPLICANT: Buck Ph.D., Jochen
; APPLICANT: Hammerling Ph.D., Ulrich
; TITLE OF INVENTION: ISOLATION, PURIFICATION AND CLONING OF
; TITLE OF INVENTION: RETINOL DEHYDRATASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,481
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/017,178
; FILING DATE: 09-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/771
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600


```
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 350 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-852-481-2

Query Match      0.8%; Score 8; DB 2; Length 350;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 668 LADGEHFI 675
DB 303 LADGEHFI 310

RESULT 5
US-09-328-352-4367
; Sequence 4367, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4367
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4367

Query Match      0.8%; Score 8; DB 4; Length 364;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 997 LSLITAIK 1004
DB 164 LSLITAIK 171

RESULT 6
US-09-248-796A-19382
; Sequence 19382, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19382
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19382

Query Match      0.8%; Score 8; DB 4; Length 373;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 LEKELDLL 802
      |||||||
```

```
DB 38 LEKELDLL 45

RESULT 7
US-08-531-601-1
; Sequence 1, Application US/08531601
; Patent No. 568684
; GENERAL INFORMATION:
; APPLICANT: YOSHIGI, NAOHIRO
; APPLICANT: MAEBA, HIDEO
; APPLICANT: OKADA, YUKIO
; TITLE OF INVENTION: RECOMBINANT BETA-AMYLASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/531,601
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP HEI 6-233086
; FILING DATE: 28-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 08/531,601
; REFERENCE/DOCKET NUMBER: 2589-031-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 531 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-531-601-1

Query Match      0.8%; Score 8; DB 1; Length 531;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVECKGP 811
      |||||||
DB 51 GLVECKGP 58

RESULT 8
US-08-859-032-1
; Sequence 1, Application US/08859032
; Patent No. 5863784
; GENERAL INFORMATION:
; APPLICANT: YOSHIGI, NAOHIRO
; APPLICANT: MAEBA, HIDEO
; APPLICANT: OKADA, YUKIO
; TITLE OF INVENTION: RECOMBINANT BETA-AMYLASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
```

```

; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/859,032
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/531,601
; FILING DATE: 21-SEP-1995
; APPLICATION NUMBER: JP HEI 6-233086
; FILING DATE: 28-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 08/531,601
; REFERENCE/DOCKET NUMBER: 2589-031-0
; TELEPHONE: (703) 413-3000
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-859-032-1

Query Match 0.8%; Score 8; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 51 GLVEGKGP 58

RESULT 9
US-08-737-597-10
; Sequence 10, Application US/08737597
; Patent No. 572657
; GENERAL INFORMATION:
; APPLICANT: YOSHIGI, NAOHICO
; TITLE OF INVENTION: BARLEY BETA AMYLASE STRUCTURAL GENE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,597
; FILING DATE: 25-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/00799
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7/92004

```

```

; FILING DATE: 25-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-044-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 535 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-737-597-10

Query Match 0.8%; Score 8; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62

RESULT 10
US-09-227-421-2
; Sequence 2, Application US/09227421
; Patent No. 6559357
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; TITLE OF INVENTION: Methods for Altering Organ Mass, Controlling Fertility
; TITLE OF INVENTION: and Enhancing Asexual Reproduction in Plants
; FILE REFERENCE: 023070-090700PC
; CURRENT APPLICATION NUMBER: US/09/227,421
; CURRENT FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: US 09/227,421
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-09-227-421-2

Query Match 0.8%; Score 8; DB 4; Length 555;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVLTL 12
DB 536 LKTVLTL 543

RESULT 11
US-09-479-855-2
; Sequence 2, Application US/09479855
; Patent No. 6639128
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; TITLE OF INVENTION: Methods for Altering Organ Mass, Controlling Fertility
; TITLE OF INVENTION: and Enhancing Asexual Reproduction in Plants
; FILE REFERENCE: 023070-090720US
; CURRENT APPLICATION NUMBER: US/09/479,855
; CURRENT FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1

```

```
; SEQ ID NO 2
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; OTHER INFORMATION: AINTEGUMENTA (ANT)
US-09-479-855-2

Query Match      0.8%; Score 8; DB 4; Length 555;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 LKTVLTL 12
DB      536 LKTVLTL 543

RESULT 12
US-09-016-000-2
; Sequence 2, Application US/09016000
; Patent No. 5962232
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Akerblom, Ingrid E.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl G.
; TITLE OF INVENTION: PROTEIN KINASE MOLECULES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,000
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0465 US
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HEARNOT01
; CLONE: 307624
US-09-016-000-2

Query Match      0.8%; Score 8; DB 2; Length 688;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      944 SEKSSESSE 951

; SEQ ID NO 2
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; OTHER INFORMATION: AINTEGUMENTA (ANT)
US-09-479-855-2

Query Match      0.8%; Score 8; DB 4; Length 555;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 LKTVLTL 12
DB      536 LKTVLTL 543

RESULT 13
US-08-560-005-2
; Sequence 2, Application US/08560005
; Patent No. 6001354
; GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Jefferson, Anne Bennett
; APPLICANT: Majerus, Philip W.
; TITLE OF INVENTION: No. 6001354el Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,005
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-560-005-2

Query Match      0.8%; Score 8; DB 3; Length 976;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      901 SKLSQLTS 908
DB      50 SKLSQLTS 57

RESULT 14
US-09-195-868-14
; Sequence 14, Application US/09195868
; Patent No. 6090621
; GENERAL INFORMATION:
; APPLICANT: KAVANAUGH MD, MICHAEL
; APPLICANT: POT PH D, DAVID
; APPLICANT: WILLIAMS MDEHD, LEWIS T.
; TITLE OF INVENTION: SIGNALING INOSITOL POLYPHOSPHATE
; TITLE OF INVENTION: 5-PHOSPHATASES (SIPs)
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 HORTON STREET
; CITY: EMERYVILLE
; STATE: CA
; COUNTRY: USA
```

```
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; CLASSIFICATION:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FIRESTONE, LEIGH H.
; REGISTRATION NUMBER: 36,831
; REFERENCE/DOCKET NUMBER: 1182.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-923-2707
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-195-868-14

Query Match 0.8%; Score 8; DB 3; Length 976;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 SKLSQ LTS 908
DB 50 SKLSQ LTS 57

RESULT 15
US-09-418-540-2
; Sequence 2, Application US/09418540
; Patent No. 6296848
; GENERAL INFORMATION:
; APPLICANT: Pot. David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Jefferson, Anne Bennett
; APPLICANT: Majerus, Philip W.
; TITLE OF INVENTION: No. 6296848el Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/418,540
; FILING DATE: 14-OCT-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/560,005
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FILING DATE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-969-528-2

Query Match 0.8%; Score 8; DB 4; Length 976;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 SKLSQ LTS 908
DB 50 SKLSQ LTS 57

RESULT 16
US-09-969-528-2
; Sequence 2, Application US/09969528
; Patent No. 6472197
; GENERAL INFORMATION:
; APPLICANT: Pot. David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Jefferson, Anne Bennett
; APPLICANT: Majerus, Philip W.
; TITLE OF INVENTION: No. 6472197el Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/969,528
; FILING DATE: 01-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/560,005
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FILING DATE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-969-528-2

Query Match 0.8%; Score 8; DB 4; Length 976;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 SKLSQ LTS 908
DB 50 SKLSQ LTS 57
```

RESULT 17
US-08-664-962B-8
; Sequence 8, Application US/08664962B
; Patent No. 6218162
; GENERAL INFORMATION:
; APPLICANT: Krystal, Gerald
; TITLE OF INVENTION: SH2-CONTAINING INOSITOL-PHOSPHATASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MERCHANT & GOULD
; STREET: 3100 No. 6218162 West Center, 90 South Seventh Street
; CITY: Minneapolis
; STATE: Minnesota
; COUNTRY: U.S.A.
; ZIP: 55402-4131
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/664,962B
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueller, Douglas P.
; REGISTRATION NUMBER: 30,300
; REFERENCE/DOCKET NUMBER: M&G 7933.49-US-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-664-962B-8

Query Match 0.8%; Score 8; DB 3; Length 1187;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 SKLSQTS 908
Db 261 SKLSQTS 268

RESULT 18
US-09-311-743-8
; Sequence 8, Application US/09311743
; Patent No. 6238903
; GENERAL INFORMATION:
; APPLICANT: Krystal, Gerald
; TITLE OF INVENTION: SH2-CONTAINING INOSITOL-PHOSPHATASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/311,743
; FILING DATE: 14-May-1999

CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7771-32
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-364-7311
; TELEFAX: 416-361-1398
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-311-743-8

Query Match 0.8%; Score 8; DB 3; Length 1187;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 SKLSQTS 908
Db 261 SKLSQTS 268

RESULT 19
US-09-195-868-15
; Sequence 15, Application US/09195868
; Patent No. 6090621
; GENERAL INFORMATION:
; APPLICANT: KAVANAUGH MD, MICHAEL
; APPLICANT: POT PH.D., DAVID
; APPLICANT: WILLIAMS MDPHD, LEWIS T.
; TITLE OF INVENTION: SIGNALING INOSITOL POLYPHOSPHATE
; TITLE OF INVENTION: 5-PHOSPHATASES (SIPs)
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 HORTON STREET
; CITY: EMERYVILLE
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; CLASSIFICATION:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FIRESTONE, LEIGH H.
; REGISTRATION NUMBER: 36,831
; REFERENCE/DOCKET NUMBER: 1182.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-923-2707
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1189 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-195-868-15

Query Match 0.8%; Score 8; DB 3; Length 1189;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 SKLSQTS 908
Db 263 SKLSQTS 270

RESULT 20

US-09-195-868-28
; Sequence 28, Application US/09195868
; Patent No. 6090621
; GENERAL INFORMATION:
; APPLICANT: KAVANAUGH MD, MICHAEL
; APPLICANT: POT PH.D., DAVID
; APPLICANT: WILLIAMS MDPHD, LEWIS T.
; TITLE OF INVENTION: SIGNALING INOSITOL POLYPHOSPHATE
; TITLE OF INVENTION: 5-PHOSPHATASES (SIPs)
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 HORTON STREET
; CITY: EMERYVILLE
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; CLASSIFICATION:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FIRESTONE, LEIGH H.
; REGISTRATION NUMBER: 36,831
; REFERENCE/DOCKET NUMBER: 1182.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-923-2707
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1229 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-195-868-28

Query Match 0.8%; Score 8; DB 3; Length 1229;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 SKLSQTS 908
Db 303 SKLSQTS 310

RESULT 21

US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21

; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match 0.8%; Score 8; DB 4; Length 3878;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLEKDKL 523
Db 357 TLEKDKL 364

RESULT 22

US-09-057-052-8
; Sequence 8, Application US/09057052
; Patent No. 6331422
; GENERAL INFORMATION:
; APPLICANT: Hubbell, Jeffrey A.
; APPLICANT: Schense, Jason
; TITLE OF INVENTION: Enzyme-Mediated Modification of Fibrin for
; TITLE OF INVENTION: Tissue Engineering
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Vinson & Elkins L.L.P.
; STREET: 600 Congress Avenue, Suite 2700
; CITY: Austin
; STATE: Texas
; COUNTRY: US
; ZIP: 78701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/057,052
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,143
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mayfield, Denise L.
; REGISTRATION NUMBER: 33,732
; REFERENCE/DOCKET NUMBER: CAL430.23000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 495-8400
; TELEFAX: (512) 495-8612
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-057-052-8

Query Match 0.7%; Score 7; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 DTIGEGQ 861
Db 4 DTIGEGQ 10

RESULT 23

US-09-057-052-12
; Sequence 12, Application US/09057052

Patent No. 6331422
GENERAL INFORMATION:
APPLICANT: Hubbell, Jeffrey A.
APPLICANT: Schense, Jason
TITLE OF INVENTION: Enzyme-Mediated Modification of Fibrin for
Tissue Engineering
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Vinson & Elkins L.L.P.
STREET: 600 Congress Avenue, Suite 2700
CITY: Austin
STATE: Texas
COUNTRY: US
ZIP: 78701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/057,052
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/042,143
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mayfield, Denise L.
REGISTRATION NUMBER: 33,732
REFERENCE/DOCKET NUMBER: CAL430.23000
TELEPHONE: (512) 495-8400
TELEFAX: (512) 495-8612
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-057-052-12

Query Match 0.7%; Score 7; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 DTIGEGQ 861
Db 4 DTIGEGQ 10

RESULT 24
US-09-695-466-8
Sequence 8, Application US/09695466
Patent No. 6607740
GENERAL INFORMATION:
APPLICANT: Hubbell, Jeffrey A.
APPLICANT: Schense, Jason
TITLE OF INVENTION: Enzyme-Mediated Modification of Fibrin for
Tissue Engineering
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Vinson & Elkins L.L.P.
STREET: 600 Congress Avenue, Suite 2700
CITY: Austin
STATE: Texas
COUNTRY: US
ZIP: 78701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/695,466
FILING DATE: 24-Oct-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/057,052
FILING DATE: <Unknown>
APPLICATION NUMBER: US 60/042,143
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mayfield, Denise L.
REGISTRATION NUMBER: 33,732
REFERENCE/DOCKET NUMBER: CAL430.23000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 495-8400
TELEFAX: (512) 495-8612
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-695-466-8

Query Match 0.7%; Score 7; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 DTIGEGQ 861
Db 4 DTIGEGQ 10

RESULT 25
US-09-695-466-12
Sequence 12, Application US/09695466
Patent No. 6607740
GENERAL INFORMATION:
APPLICANT: Hubbell, Jeffrey A.
APPLICANT: Schense, Jason
TITLE OF INVENTION: Enzyme-Mediated Modification of Fibrin for
Tissue Engineering
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Vinson & Elkins L.L.P.
STREET: 600 Congress Avenue, Suite 2700
CITY: Austin
STATE: Texas
COUNTRY: US
ZIP: 78701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/695,466
FILING DATE: 24-Oct-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/057,052
FILING DATE: <Unknown>
APPLICATION NUMBER: US 60/042,143
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mayfield, Denise L.
REGISTRATION NUMBER: 33,732
REFERENCE/DOCKET NUMBER: CAL430.23000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 495-8400
TELEFAX: (512) 495-8612
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-695-466-12

Query Match          0.7%; Score 7; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 855 DTIGEGQ 861
Db 4 DTIGEGQ 10

RESULT 26
US-09-270-767-60016
; Sequence 60016, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60016
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-60016

Query Match          0.7%; Score 7; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 825 KTLPLPLP 831
Db 12 KTLPLPLP 18

RESULT 27
US-09-270-767-44565
; Sequence 44565, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44565
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-44565

Query Match          0.7%; Score 7; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 825 KTLPLPLP 831
Db 24 KTLPLPLP 30

RESULT 28
US-09-270-767-61674
; Sequence 61674, Application US/09270767
```

```
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61674
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-61674

Query Match          0.7%; Score 7; DB 4; Length 39;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 ASSSISG 44
Db 11 ASSSISG 17

RESULT 29
US-09-621-976-6627
; Sequence 6627, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.Pm
; SEQ ID NO 6627
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6627

Query Match          0.7%; Score 7; DB 4; Length 60;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 GKGPLAQ 814
Db 28 GKGPLAQ 34

RESULT 30
US-09-621-976-6631
; Sequence 6631, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.Pm
; SEQ ID NO 6631
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6631
```



```
Query Match      0.7%; Score 7; DB 4; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      808 GKGPLAQ 814
DB      28 GKGPLAQ 34

RESULT 31
US-09-248-796A-20152
; Sequence 20152, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20152
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20152

Query Match      0.7%; Score 7; DB 4; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      994 VALLSLI 1000
DB      12 VALLSLI 18

RESULT 32
US-09-513-999C-5429
; Sequence 5429, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1998-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5429
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: UNSURE
; LOCATION: 73
; OTHER INFORMATION: Xaa=Glu or Gly or Lys or Arg
US-09-513-999C-5429

Query Match      0.7%; Score 7; DB 4; Length 77;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      995 ALLSLIT 1001
DB      11 ALLSLIT 1001

Query Match      0.7%; Score 7; DB 4; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      808 GKGPLAQ 814
DB      28 GKGPLAQ 34

RESULT 33
US-09-621-976-6626
; Sequence 6626, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6626
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6626

Query Match      0.7%; Score 7; DB 4; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      808 GKGPLAQ 814
DB      49 GKGPLAQ 55

RESULT 34
US-09-621-976-6628
; Sequence 6628, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6628
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6628

Query Match      0.7%; Score 7; DB 4; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      808 GKGPLAQ 814
DB      49 GKGPLAQ 55

RESULT 35
US-09-621-976-6629
; Sequence 6629, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
```

; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6629
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6629

Query Match 0.7%; Score 7; DB 4; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 KGKPLAQ 814
|||||
Db 49 KGKPLAQ 55

RESULT 36

US-09-621-976-6630
; Sequence 6630, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6630
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 36
; OTHER INFORMATION: Xaa = Asp, Glu, Val
US-09-621-976-6630

Query Match 0.7%; Score 7; DB 4; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 KGKPLAQ 814
|||||
Db 49 KGKPLAQ 55

RESULT 37

US-09-248-796A-19130
; Sequence 19130, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19130
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19130

Query Match 0.7%; Score 7; DB 4; Length 85;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 769 QLLERSI 775
|||||
Db 77 QLLERSI 83

RESULT 38

US-09-248-796A-16501
; Sequence 16501, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16501
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16501

Query Match 0.7%; Score 7; DB 4; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 692 LKLI EGL 698
|||||
Db 12 LKLI EGL 18

RESULT 39

US-09-270-767-40137
; Sequence 40137, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1998-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 40137
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-40137

Query Match 0.7%; Score 7; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 225 LLKKVTI 231
|||||
Db 25 LLKKVTI 31

RESULT 40

US-09-270-767-55353
; Sequence 55353, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55353
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-55353

Query Match 0.7%; Score 7; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 LKKVTI 231
|||||
Db 25 LKKVTI 31

RESULT 41
US-09-270-767-44382
; Sequence 44382, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44382
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-44382

Query Match 0.7%; Score 7; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 KDLAKG 90
|||||
Db 26 KDLAKG 32

RESULT 42
US-09-902-540-11181
; Sequence 11181, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 11181
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-11181

Query Match 0.7%; Score 7; DB 4; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 ALVTCKA 764
|||||
Db 69 ALVTCKA 75
RESULT 43
US-09-513-999C-5997
; Sequence 5997, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5997
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-5997

Query Match 0.7%; Score 7; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 LVGLSKS 152
|||||
Db 10 LVGLSKS 16

RESULT 44
US-09-270-767-44049
; Sequence 44049, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44049
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-44049

Query Match 0.7%; Score 7; DB 4; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 946 KSSSES 952
|||||
Db 90 KSSSES 96

RESULT 45
US-09-248-796A-20841
; Sequence 20841, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132

;; CURRENT APPLICATION NUMBER: US/09/248,796A
;; CURRENT FILING DATE: 1999-02-12
;; PRIOR APPLICATION NUMBER: US 60/074,725
;; PRIOR FILING DATE: 1998-02-13
;; PRIOR APPLICATION NUMBER: US 60/096,409
;; PRIOR FILING DATE: 1998-08-13
;; NUMBER OF SEQ ID NOS: 28208
;; SEQ ID NO 20841
;; LENGTH: 111
;; TYPE: PRT
;; ORGANISM: Candida albicans
US-09-248-796A-20841

Query Match 0.7%; Score 7; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 996 LLSLITA 1002
|||||||
Db 32 LLSLITA 38

RESULT 46
US-09-071-035-206
; Sequence 206, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 206:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-206

Query Match 0.7%; Score 7; DB 4; Length 120;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 975 LPRTGSK 981
|||||||
Db 85 LPRTGSK 91

US-09-071-035-206
; Sequence 206, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 206:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-206

Query Match 0.7%; Score 7; DB 4; Length 120;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 975 LPRTGSK 981
|||||||
Db 85 LPRTGSK 91

RESULT 47
US-09-071-035-210
; Sequence 210, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 210:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-210

Query Match 0.7%; Score 7; DB 4; Length 120;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 975 LPRTGSK 981
|||||||
Db 85 LPRTGSK 91

RESULT 48
US-09-252-991A-28777
; Sequence 28777, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28777
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-28777

Query Match 0.7%; Score 7; DB 4; Length 121;
Best Local Similarity 100.0%; Pred. No. 2e+02; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 860 GOKDAYG 866
DB 97 GOKDAYG 103

RESULT 49

US-09-270-767-33193
; Sequence 33193, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33193
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-33193

Query Match 0.7%; Score 7; DB 4; Length 123;
Best Local Similarity 100.0%; Pred. No. 2e+02; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 930 AAABEOL 936
DB 102 AAABEOL 108

RESULT 50

US-09-270-767-48410
; Sequence 48410, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48410
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-48410

Query Match 0.7%; Score 7; DB 4; Length 123;
Best Local Similarity 100.0%; Pred. No. 2e+02; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 930 AAABEOL 936
DB 102 AAABEOL 108

RESULT 51

US-09-270-767-57861
; Sequence 57861, Application US/09270767
; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57861
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-57861

Query Match 0.7%; Score 7; DB 4; Length 136;
Best Local Similarity 100.0%; Pred. No. 2.2e+02; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 701 STLQTT 707
DB 10 STLQTT 16

RESULT 52

US-09-134-000C-6044
; Sequence 6044, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6044
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6044

Query Match 0.7%; Score 7; DB 4; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.3e+02; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 975 LPRTGSK 981
DB 108 LPRTGSK 114

RESULT 53

US-08-858-207A-523
; Sequence 523, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgeson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 523:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
; US-08-858-207A-523

Query Match          0.7%; Score 7; DB 3; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 EIKGAP 509
   |||||
Db 45 EIKGAF 51

RESULT 54
US-09-134-000C-6221
; Sequence 6221, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6221
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-09-134-000C-6221

Query Match          0.7%; Score 7; DB 4; Length 147;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 460 FONKTLR 466
   |||||
Db 5 FONKTLR 11

RESULT 55
US-08-679-493A-203
; Sequence 203, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
```

```
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; CURRENT FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 203
; LENGTH: 152
; TYPE: PRT
; ORGANISM: guineapig
; US-08-679-493A-203

Query Match          0.7%; Score 7; DB 3; Length 152;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 803 TGLVEGK 809
   |||||
Db 35 TGLVEGK 41

RESULT 56
US-09-252-991A-24022
; Sequence 24022, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24022
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-24022

Query Match          0.7%; Score 7; DB 4; Length 161;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 IAKAEKA 758
   |||||
Db 149 IAKAEKA 155

RESULT 57
US-09-538-092-450
; Sequence 450, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Ioic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
```

```
; SEQ ID NO 450
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YJL124C
US-09-538-092-450

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 172;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 952 SANSKDR 958
Db 2 SANSKDR 8

RESULT 58
US-09-270-767-38641
; Sequence 38641, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38641
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-38641

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 176;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 TKKATKN 767
Db 161 TKKATKN 167

RESULT 59
US-09-270-767-53858
; Sequence 53858, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53858
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-53858

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 176;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 TKKATKN 767
Db 161 TKKATKN 167

RESULT 60
US-09-893-737-22
; Sequence 22, Application US/09893737
; Patent No. 6822082
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-893-737-22

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 176;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 VKRLEKE 798
Db 114 VKRLEKE 120

RESULT 61
US-09-270-767-35060
; Sequence 35060, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35060
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-35060

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 180;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 575 AFLSNRL 581
Db 48 AFLSNRL 54

RESULT 62
US-09-270-767-50277
; Sequence 50277, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
```

```
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 50277
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-50277

Query Match          0.7%; Score 7; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 575 AFLSNRL 581
Db 48 AFLSNRL 54

RESULT 63
US-09-270-767-61846
; Sequence 61846, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61846
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-61846

Query Match          0.7%; Score 7; DB 4; Length 181;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 895 IKTILNS 901
Db 172 IKTILNS 178

RESULT 64
US-09-902-540-15188
; Sequence 15188, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wisland, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 15188
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-15188

Query Match          0.7%; Score 7; DB 4; Length 190;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 7 TVALTLT 13
Db 23 TVALTLT 29
```

```
RESULT 65
US-09-270-767-46119
; Sequence 46119, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46119
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46119
```

```
Query Match          0.7%; Score 7; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 38 ASSSISG 44
Db 11 ASSSISG 17
```

```
RESULT 66
US-09-248-796A-15823
; Sequence 15823, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15823
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-15823
```

```
Query Match          0.7%; Score 7; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 561 LIFMGSK 567
Db 146 LIFMGSK 152
```

```
RESULT 67
US-09-305-984-18
; Sequence 18, Application US/09305984B
; Patent No. 6331407
; GENERAL INFORMATION:
; APPLICANT: No. 6331407ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
```


; CURRENT APPLICATION NUMBER: US/09/305,984B
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-305-984-18

Query Match 0.7%; Score 7; DB 3; Length 215;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522
Db 199 TLELKD 205

RESULT 68
US-09-305-984-64
; Sequence 64, Application US/09305984B
; Patent No. 6331407
; GENERAL INFORMATION:
; APPLICANT: No. 6331407ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/305,984B
; EARLIER FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptococcus Pneumoniae
US-09-305-984-64

Query Match 0.7%; Score 7; DB 3; Length 215;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522
Db 199 TLELKD 205

RESULT 69
US-09-073-541A-18
; Sequence 18, Application US/09073541A
; Patent No. 6448224
; GENERAL INFORMATION:
; APPLICANT: No. 6448224ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016
; CURRENT APPLICATION NUMBER: US/09/073,541A
; CURRENT FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-073-541A-18

Query Match 0.7%; Score 7; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522
Db 199 TLELKD 205

RESULT 70
US-09-493-940-18
; Sequence 18, Application US/09493940
; Patent No. 6630583
; GENERAL INFORMATION:
; APPLICANT: No. 6630583ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/493,940
; CURRENT FILING DATE: 2000-01-28
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-493-940-18

Query Match 0.7%; Score 7; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522
Db 199 TLELKD 205

RESULT 71
US-09-493-940-64
; Sequence 64, Application US/09493940
; Patent No. 6630583
; GENERAL INFORMATION:
; APPLICANT: No. 6630583ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/493,940
; CURRENT FILING DATE: 2000-01-28
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptococcus Pneumoniae
US-09-493-940-64

Query Match 0.7%; Score 7; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522
Db 199 TLELKD 205

```
RESULT 72
US-09-583-110-5320
; Sequence 5320, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5320
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5320

Query Match          0.7%; Score 7; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      516 TLELKDK 522
      |||||
Db      199 TLELKDK 205

RESULT 73
US-09-107-433-4810
; Sequence 4810, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4810:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5320

Query Match          0.7%; Score 7; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      516 TLELKDK 522
      |||||
Db      199 TLELKDK 205
```

```
; LENGTH: 219 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...219
; SEQUENCE DESCRIPTION: SEQ ID NO: 4810:
US-09-107-433-4810

Query Match          0.7%; Score 7; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      516 TLELKDK 522
      |||||
Db      203 TLELKDK 209

RESULT 74
US-09-305-984-72
; Sequence 72, Application US/09305984B
; Patent No. 6331407
; GENERAL INFORMATION:
; APPLICANT: No. 6331407ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/305,984B
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-305-984-72

Query Match          0.7%; Score 7; DB 3; Length 224;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      516 TLELKDK 522
      |||||
Db      208 TLELKDK 214

RESULT 75
US-09-305-984-74
; Sequence 74, Application US/09305984B
; Patent No. 6331407
; GENERAL INFORMATION:
; APPLICANT: No. 6331407ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/305,984B
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 74
; LENGTH: 224
```

;
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-305-984-74

Query Match 0.7%; Score 7; DB 3; Length 224;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKD 522
Db 208 TLELKD 214

RESULT 76
US-09-305-984-76
; Sequence 76, Application US/09305984B
; Patent No. 6331407
; GENERAL INFORMATION:
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/305,984B
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 76
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-305-984-76

Query Match 0.7%; Score 7; DB 3; Length 224;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKD 522
Db 208 TLELKD 214

RESULT 77
US-09-493-940-72
; Sequence 72, Application US/09493940
; Patent No. 6630583
; GENERAL INFORMATION:
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/493,940
; CURRENT FILING DATE: 2000-01-28
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-493-940-72

Query Match 0.7%; Score 7; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKD 522

Db 208 TLELKD 214

RESULT 78
US-09-493-940-74
; Sequence 74, Application US/09493940
; Patent No. 6630583
; GENERAL INFORMATION:
; APPLICANT: No. 6630583ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/493,940
; CURRENT FILING DATE: 2000-01-28
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 74
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-493-940-74

Query Match 0.7%; Score 7; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKD 522
Db 208 TLELKD 214

RESULT 79
US-09-493-940-76
; Sequence 76, Application US/09493940
; Patent No. 6630583
; GENERAL INFORMATION:
; APPLICANT: No. 6630583ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/493,940
; CURRENT FILING DATE: 2000-01-28
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 76
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-493-940-76

Query Match 0.7%; Score 7; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKD 522
Db 208 TLELKD 214

RESULT 80
US-09-902-540-11067
; Sequence 11067, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:

```
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 11067
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-11067

Query Match          0.7%; Score 7; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 143 GNTLVGL 149
Db 96 GNTLVGL 102

RESULT 81
US-09-198-452A-399
; Sequence 399, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 399
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-399

Query Match          0.7%; Score 7; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 595 IPVQAPS 601
Db 112 IPVQAPS 118

RESULT 82
US-09-438-185A-384
; Sequence 384, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kalman, Sue
; APPLICANT: Davis, Ronald
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-000411US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 384
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; OTHER INFORMATION: CPn0382
US-09-438-185A-384

Query Match          0.7%; Score 7; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 595 IPVQAPS 601
Db 112 IPVQAPS 118

RESULT 83
US-08-867-087B-11
; Sequence 11, Application US/08867087B
; Patent No. 5990386
; GENERAL INFORMATION:
; APPLICANT: An, Gynheung
; TITLE OF INVENTION: GENES CONTROLLING FLORAL DEVELOPMENT
; TITLE OF INVENTION: AND APICAL DOMINANCE IN PLANTS
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,087B
; FILING DATE: June 2, 1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/323,449
; FILING DATE: October 14, 1994
; APPLICATION NUMBER: U.S. 08/485,981
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Alan. E.
; REGISTRATION NUMBER: 35,123
; REFERENCE/DOCKET NUMBER: 4630-47071
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acid residues
; TYPE: amino acid
; TOPOLOGY: linear
US-08-867-087B-11

Query Match          0.7%; Score 7; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 704 RQTTQTQ 710
Db 194 RQTTQTQ 200
```

RESULT 84
US-09-107-532A-5876
; Sequence 5876, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5876:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...241
; SEQUENCE DESCRIPTION: SEQ ID NO: 5876:
US-09-107-532A-5876
Query Match 0.7%; Score 7; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 793 KRLKEL 799
Db 126 KRLKEL 132
US-09-107-532A-5876
RESULT 85
US-09-252-991A-21887
; Sequence 21887, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32999
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32999
Query Match 0.7%; Score 7; DB 4; Length 248;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 717 AGKALLS 723

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21887
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21887
Query Match 0.7%; Score 7; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 882 LAVATLA 888
Db 27 LAVATLA 33
US-09-328-352-4519
; Sequence 4519, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4519
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4519
Query Match 0.7%; Score 7; DB 4; Length 243;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 254 LPESLET 260
Db 63 LPESLET 69
US-09-252-991A-32999
; Sequence 32999, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32999
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32999
Query Match 0.7%; Score 7; DB 4; Length 248;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 717 AGKALLS 723

```
Db      118 AGKALLS 124
|||||
RESULT 88
US-09-270-767-40859
; Sequence 40859, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40859
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-40859
Query Match      0.7%; Score 7; DB 4; Length 250;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      256 ESLETIS 262
      |||||
Db      65 ESLETIS 71

RESULT 89
US-09-270-767-56075
; Sequence 56075, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56075
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-56075
Query Match      0.7%; Score 7; DB 4; Length 250;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      256 ESLETIS 262
      |||||
Db      65 ESLETIS 71

RESULT 90
US-09-248-796A-17359
; Sequence 17359, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17359
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17359
Query Match      0.7%; Score 7; DB 4; Length 256;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      684 TIVDEK 690
      |||||
Db      67 TIVDEK 73

RESULT 91
US-09-489-039A-7271
; Sequence 7271, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7271
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7271
Query Match      0.7%; Score 7; DB 4; Length 259;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      750 KAIKAE 756
      |||||
Db      59 KAIKAE 65

RESULT 92
US-09-489-039A-11968
; Sequence 11968, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11968
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11968
Query Match      0.7%; Score 7; DB 4; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      378 NPSGLAT 384
      |||||
Db      215 NPSGLAT 221
```

```
RESULT 93
US-09-270-767-42555
; Sequence 42555, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 42555
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-42555

Query Match      0.7%; Score 7; DB 4; Length 268;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 701 STLQTT 707
   |||||
Db 142 STLQTT 148

RESULT 94
US-09-248-796A-19421
; Sequence 19421, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19421
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19421

Query Match      0.7%; Score 7; DB 4; Length 270;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 NSKLSQL 906
   |||||
Db 137 NSKLSQL 143

RESULT 95
US-09-489-039A-10410
; Sequence 10410, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
```

```
; SEQ ID NO 10410
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10410

Query Match      0.7%; Score 7; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 992 TSVALLS 998
   |||||
Db 18 TSVALLS 24

RESULT 96
US-08-741-437-1
; Sequence 1, Application US/08741437
; Patent No. 5843665
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/741,437
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0148 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555
; TELEFAX: (415) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
US-08-741-437-1

Query Match      0.7%; Score 7; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 757 KALVTKK 763
   |||||
Db 228 KALVTKK 234

RESULT 97
US-09-134-593-1
; Sequence 1, Application US/09134593
```

Patent No. 5981232
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,593
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/741,437
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0148 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555
; TELEFAX: (415) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
US-09-134-593-1

Query Match 0.7%; Score 7; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 757 KALVTKK 763
Db 228 KALVTKK 234

RESULT 98
US-09-248-796A-18420
; Sequence 18420, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18420
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18420

Query Match 0.7%; Score 7; DB 4; Length 289;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 763 KATKNGQ 769
Db 195 KATKNGQ 201

RESULT 99
US-09-538-092-1335
; Sequence 1335, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 1335
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q15181
US-09-538-092-1335

Query Match 0.7%; Score 7; DB 4; Length 289;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 757 KALVTKK 763
Db 228 KALVTKK 234

RESULT 100
US-09-535-521-2
; Sequence 2, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
; APPLICANT: Weber, Eric R.
; APPLICANT: McCall, Catherine A.
; TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IGE RECEPTOR (CANINE CD23)
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
; FILE REFERENCE: AL-5
; CURRENT APPLICATION NUMBER: US/09/535,521
; CURRENT FILING DATE: 2000-03-24
; EARLIER APPLICATION NUMBER: 60/125,913
; EARLIER FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-535-521-2

Query Match 0.7%; Score 7; DB 4; Length 292;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 627 LKQLEVA 633

Db |||||||
 57 LKQLEVA 63

Search completed: August 28, 2005, 11:17:45
Job time : 46 secs

THIS PAGE BLANK (USPTO)